
UNIVERSIDADE DE LISBOA

**Faculdade de Medicina Veterinária**

# ASPECTS OF MOLECULAR ECOLOGY OF CARNIVORE VIRUSES: *SAPOVIRUS* AND *CORONAVIRUS*

JOANA FORTUNA DOS REMÉDIOS

CONSTITUIÇÃO DO JÚRI

Doutor Virgílio da Silva Almeida

Doutora Ana Isabel Simões Pereira Duarte

Doutor João Nestor das Chagas e Silva

ORIENTADORA

Doutora Marion L East

CO-ORIENTADOR

Doutor João Nestor das Chagas e Silva

2018

LISBOA


UNIVERSIDADE DE LISBOA

**Faculdade de Medicina Veterinária**

# ASPECTS OF MOLECULAR ECOLOGY OF CARNIVORE VIRUSES: *SAPOVIRUS* AND *CORONAVIRUS*

JOANA FORTUNA DOS REMÉDIOS

DISSERTAÇÃO DE MESTRADO INTEGRADO EM MEDICINA VETERINÁRIA

| CONSTITUIÇÃO DO JÚRI | ORIENTADORA |
|---|---|
| Doutor Virgílio da Silva Almeida | Doutora Marion L East |
| Doutora Ana Isabel Simões Pereira Duarte | CO-ORIENTADOR |
| Doutor João Nestor das Chagas e Silva | Doutor João Nestor das Chagas e Silva |

2018

LISBOA

**DECLARAÇÃO**

**Nome:** Joana Fortuna dos Remédios

**Título:** Dissertação
"Aspects of molecular ecology of carnivore viruses: *Sapovirus* and *Coronavirus*"

**Orientadora:** Doutora Marion L East

**Co-orientador:** Doutor João Nestor das Chagas e Silva

**Ano de conclusão:** 2018

**Designação do Mestrado ou do ramo de conhecimento do Doutoramento:**
Mestrado Integrado em Medicina Veterinária



Faculdade de Medicina Veterinária da ULisboa, 16 de Abril de 2018

**Assinatura:**_____

Aos meus avós.

Aos quatro patas da minha vida.

**Acknowledgements**

I would like to express my gratitude to Dr. Marion L East for accepting to be my supervisor during my internship in the Leibniz Institute for Zoo and Wildlife Research, presenting me the complexity and the challenges of research in wildlife and the whole world still to be discovered. I am deeply grateful for have been introduced to the spotted hyenas of the Serengeti by a person so passionate and dedicated to the species. For all of this and the guidance, even from distance, thank you.

To Dr. João Nestor das Chagas e Silva I would like to express my thanks for accepting from the first moment to be my co-supervisor, have remained despite the change in the internship area and always being available to talk and help.

I am grateful to Ximena Olarte-Castillo for sharing with me her work and guiding me through mine, to Dagmar Thierer for the crucial help at the beginning of my stay in Berlin, and to Stephan Karl for having taught and guided me through the laboratory work, for all the conversations that made the laboratory less lonely and for the help and trust in my work and in solving the problems. I would also like to thank Susana and Sónia for welcoming me so well at IZW, and Franziska with whom I shared almost all this period.

To everyone at the Ani+ clinic I want to thank you for allowing me to accompany you and learn from you.

An enormous thank you to my friends, specially Inês, Joanas (Correia, Mariano, Mendonça) and Mariana for all the pushing and motivational boosts and the hours of talk and laughter in the last months, but mostly for all these years of growing friendship.

I don't have enough words to thank Guilherme for the patience and the joy you brought me during this work. Thank you for always encouraging me to follow the paths I have chosen, for the ever-present support even in different countries and the readiness for the endless revisions of my thesis.

Lastly, a special thanks to my family for all the support throughout my life that made it possible for me to be here today, for believing in me and helping me fulfil my dreams. I thank my mother all the love, the advices and the example of strength and human being that she is. To my father I thank the teachings and the example of work and professionalism that I seek to follow. I thank my sister for a life of sharing, complicity and for being the role model I looked up to so many times just because she is my big sister, for the presence, even at a distance, for the necessary words and available help and for the laughter that lightened everything. To the cats with whom I grew up (Fifi, Dora e Ruca), and who are family, thank you for the company, for the joys and for helping me to recognize my path. To all of you who are my foundations and are always here when I need, thank you very much.

Financial support was provided by the Leibniz Institute for Zoo and Wildlife Research.

Leibniz Institute for Zoo
and Wildlife Research
IN THE FORSCHUNGSVERBUND BERLIN E.V.

**Aspetos da ecologia molecular de vírus de carnívoros: *Sapovirus* e *Coronavirus***

**Resumo:** O conhecimento atual sobre a epidemiologia de muitos agentes patogénicos em animais selvagens é limitado e pouco se sabe sobre a sua diversidade genética, a sua extensão geográfica e de espécies hospedeiras, e o seu potencial de propagação entre espécies selvagens, domésticas e humana. Esta tese, desenvolvida no Leibniz Institute for Zoo and Wildlife Research, inclui dois estudos que procuram contribuir para o aumento do conhecimento nesta área.

O primeiro estudo procurou testar a hipótese de que surtos consecutivos de infeção por *Sapovirus* (SaV) em hienas-malhadas (*Crocuta crocuta*), detetados por um estudo prévio nesta espécie no Parque Nacional do Serengeti, resultaram da emergência de estirpes antigenicamente diferentes do vírus. O RNA do vírus foi extraído de amostras fecais obtidas de três hienas infetadas e amplificado usando métodos convencionais de RT-PCR, com o objetivo de sequenciar um fragmento do genoma viral que se sabe ser importante para a determinação do tipo antigénico das estirpes de SaV. Apesar de terem sido experimentados vários conjuntos de primers, apenas foi obtida uma sequência parcial do gene-alvo de uma amostra, pelo que não foi possível determinar se os surtos de infeção por SaV entre as hienas-malhadas no Serengeti foram causados por estirpes antigenicamente distintas.

O segundo estudo visou a aminopeptidase N (APN), uma proteína conhecida como recetor celular para um grande número de alfa-coronavirus (α-CoVs). Estudos *in vitro* demonstraram que as APNs canina e felina conseguem facilitar a entrada de α-CoVs de diferentes espécies nas células destes carnívoros. Este trabalho teve por objetivo investigar a relação filogenética entre a APN de diferentes espécies de carnívoros. Uma atenção particular foi dada à região que se sabe interagir com os α-CoVs durante a sua entrada na célula, com o propósito de melhor compreender a possibilidade de α-CoVs de uma espécie hospedeira particular alargarem com sucesso o seu leque de hospedeiros. Procurou-se também a presença de isoformas da APN. A amplificação e sequenciação de nove amostras de tecidos de carnívoros selvagens foram realizadas usando métodos convencionais de RT-PCR e métodos de clonagem molecular, seguidos da análise filogenética dos resultados. Sete sequências parciais da APN foram obtidas e a sua relação filogenética correspondeu à dos seus animais de origem. No entanto, a análise da região específica onde o vírus adere revelou que as espécies das famílias Hyaenidae e Herpestidae (subordem Feliformia) eram filogeneticamente mais semelhantes a espécies da subordem Caniformia do que da subordem Feliformia. Isto sugere que α-CoVs que infetem espécies desta duas famílias possam estender a sua variedade de hospedeiros a espécies da subordem Caniformia em vez da subordem Feliformia.

Os resultados obtidos complementam a informação já existente acerca do SaV e do APN.

**Palavras-chave:** *Sapovirus*, *Coronavirus*, aminopeptidase N, carnívoros, análise filogenética, vida selvagem

**Aspects of molecular ecology of carnivore viruses: *Sapovirus* and *Coronavirus***

**Abstract:** Current knowledge on the epidemiology of many pathogens of wild animals is limited and little is known about their genetic diversity, geographic and host species range, and their potential to spread between wild, domestic and human species. This thesis, developed at the Leibniz Institute for Zoo and Wildlife Research, includes two studies that seek to contribute to increase the knowledge in this field.

The first study aimed to test the hypothesis that consecutive outbreaks of *Sapovirus* (SaV) infection in spotted hyenas (*Crocuta crocuta*), detected by a previous study on this species in the Serengeti National Park, resulted from the emergence of antigenically different strains of the virus. Virus RNA was extracted from faecal samples obtained from three infected hyenas and amplified using conventional RT-PCR methods, with the aim of sequencing a fragment of the virus genome known to be important in determining the antigenic type of SaV strains. Although a diverse set of primer pairs were tried, only a partial sequence of the targeted gene was obtained from one sample, thus it was not possible to determine if the outbreaks of SaV infection among spotted hyenas in the Serengeti were caused by distinct antigenic strains.

The second study targeted aminopeptidase N (APN), a protein known to work as the host cell receptor for a great number of alphacoronaviruses (α-CoVs). *In vitro* studies have demonstrated that canine and feline APNs can facilitate the entry of α-CoVs from different species into these carnivore's cells. This work aimed to investigate the phylogenetic relation between APN from different carnivore species. A particular focus was given to the region known to interact with α-CoVs during cell entry, with the purpose to better understand the possibility of α-CoVs from particular host species successful extending their range of hosts. The detection of isoforms of APN was also sought. The amplification and sequencing of nine tissue samples of wild carnivores was performed using conventional RT-PCR and molecular cloning methods, followed by the phylogenetic analysis of the results. Seven partial sequences of APN were obtained and their phylogenetic relation corresponded to that of their animals of origin. However, the analysis of the specific region where the virus attaches revealed that the species from the families Hyaenidae and Herpestidae (suborder Feliformia) were phylogenetically more similar to the species from Caniformia suborder rather than those from Feliformia suborder. This suggests that α-CoVs that infect the species in these two families might extend their host range to species in Caniformia rather than Feliformia suborder.

The results obtained complement the already existing information on both *Sapovirus* and APN.

**Key words:** *Sapovirus*, *Coronavirus*, aminopeptidase N, carnivores, phylogenetic analysis, wildlife

**Table of Contents:**


I. INTERNSHIP ACTIVITIES ..... 1

II. INTRODUCTION ..... 2

III. *SAPOVIRUS* IN SPOTTED HYENAS FROM SERENGETI NATIONAL PARK ..... 4

1. LITERATURE REVIEW ..... 4

1.1. *Sapovirus* ..... 4

1.1.1. *Sapovirus* Classification ..... 5

1.1.2. Genetic Diversity ..... 6

1.1.3. Life Cycle ..... 8

1.1.4. Pathophysiology ..... 9

1.1.5. Virus Detection ..... 10

1.2. Serengeti National Park ..... 11

1.2.1. Spotted Hyenas (*Crocuta crocuta*) of Serengeti National Park ..... 13

1.2.1.1. Feeding Behaviour ..... 14

1.2.1.2. Territory ..... 15

1.2.2. Serengeti National Park as a Research Subject ..... 15

1.2.2.1. *Sapovirus* in Serengeti National Park ..... 16

1.3. Aim of the Study ..... 18

2. MATERIAL AND METHODS ..... 19

2.1. Samples ..... 19

2.2. RNA Extraction ..... 19

2.3. RT-PCR, Electrophoresis and Sequencing ..... 19

2.4. Phylogenetic Analysis ..... 22

3. RESULTS ..... 24

4. DISCUSSION ..... 26

4.1. *Sapovirus* Sequences ..... 26

4.2. Spotted Hyena' *Sapovirus* and Phylogeny ..... 27

IV. *ALPHACORONAVIRUS* RECEPTOR IN WILD CARNIVORES ..... 28

1. LITERARTURE REVIEW ..... 28

1.1. *Coronavirus* ..... 28

1.1.1. Life cycle ..... 29

1.1.2. Pathophysiology ..... 30

1.1.3. *Alphacoronavirus* ..... 32

1.2. Aminopeptidase N ..... 32

1.2.1. Protein Synthesis ..... 33

1.2.1.1. RNA Transformation: Alternative Splicing and RNA Editing ..... 34

1.2.2. Structure ..... 35

1.2.3. Function ..... 36
1.2.3.1. APN as *Alphacoronavirus* Receptor ..... 37
1.3. Aim of the Study ..... 38
2. MATHERIAL AND METHODS ..... 40
2.1. Samples ..... 40
2.2. RNA Extraction ..... 40
2.3. RT-PCR and Electrophoresis ..... 41
2.3.1. DNA Extraction from Agarose Gel ..... 42
2.3.2. Molecular Cloning ..... 43
2.4. Sequencing ..... 44
2.5. Phylogenetic Analysis ..... 44
3. RESULTS ..... 47
3.1. APNs Sequenced ..... 47
3.2. RNA Alternative Splicing ..... 47
3.3. Phylogenetic Trees ..... 48
4. DISCUSSION ..... 52
4.1. Aminopeptidase N Sequences ..... 52
4.2. Aminopeptidase N Alternative Splicing ..... 53
4.3. Aminopeptidase N Phylogeny ..... 53
4.4. *Alphacoronavirus* Receptor's Phylogeny ..... 54
V. CONCLUSION ..... 55
VI. FUTURE PERSPECTIVES ..... 56
VII. BIBLIOGRAPHY ..... 57
VIII. ANNEX 1 – Pairs of primers used for the amplification of *Sapovirus* ..... 72
IX. ANNEX 2 – Pairs of primers used for the amplification of Aminopeptidase N ..... 73

**List of figures:**

Figure 1. *Sapovirus* genomic organization ..... 5
Figure 2. Possible model of bile acid mediated *Sapovirus* replication ..... 8
Figure 3. Map of Serengeti Mara Ecosystem and its protected areas ..... 11
Figure 4. Prevalence of *Sapovirus* infection in spotted hyenas from Serengeti National Park between 2001 and 2012 ..... 17
Figure 5. Picture of gel electrophoresis' result of SaV amplification with primers SaV1245Rfwd and SV-R2 ..... 24
Figure 6. Partial sequence of *Sapovirus*' VP1 encoding region found in spotted hyena (I 600) ..... 25
Figure 7. Phylogenetic relationship of the VP1's partial length genomic sequence of spotted hyena (*Crocuta crocuta*) *Sapovirus* strain (original) indicated by a black triangle ..... 25
Figure 8. Coronavirus replication model ..... 30
Figure 9. Patterns of alternative splicing ..... 35
Figure 10. Diagram of APN's receptor region for alphacoronaviruses ..... 38
Figure 11. Picture of gel electrophoresis' results with more than one band ..... 48
Figure 12. Phylogenetic relationship of aminopeptidase N's partial amino acid sequence of wild carnivores ..... 49
Figure 13. Phylogenetic relationship of the C-terminal region of aminopeptidase N with known viral receptor function for the majority of alphacoronaviruses of wild carnivores ..... 50

**List of tables:**

Table 1. Wild carnivore species that inhabit the Serengeti Ecosystem ..... 13
Table 2. Thermocycling conditions for RT-PCR ..... 20
Table 3. Primers used on *Sapovirus* amplification and sequencing ..... 21
Table 4. Thermocycling conditions for clean-up prior to sequencing ..... 21
Table 5. Thermocycling conditions for sequencing reaction ..... 22
Table 6. Tissue samples from different wild carnivores used in the study, along with their place of collection ..... 40
Table 7. Primers used on aminopeptidase N amplification and sequencing ..... 41
Table 8. Thermocycling conditions for PCR ..... 44
Table 9. Species integrated in the phylogenetic analysis ..... 46

**List of Abbreviations:**

AICc – Akaike Information Criterion, corrected
ANPEP – Alanyl aminopeptidase
α-CoV – Alphacoronavirus
aa – Amino acid
APN – Aminopeptidase N
BIC – Bayesian information criterion
BLAST – Basic Local Alignment Search Tool
β-CoV – Betacoronavirus
cAPN – Canine aminopeptidase N
CCoV – Canine coronavirus
cDNA – Complementary deoxyribonucleic acid
CoV – Coronavirus
dNTPs – Deoxynucleotide triphosphates
ddNTPs – Dideoxynucleotide triphosphates
E – Envelop protein
fAPN – Feline aminopeptidase N
FCoV – Feline enteric coronavirus
FIPV – Feline infectious peritonitis virus
γ-CoV – Gammacoronavirus
G – Genogroup
hAPN – Human aminopeptidase N
HCoV-229E – Human coronavírus strain 229E
HCoV-HKU1 – Human coronavírus strain HKU1
HCoV-NL63 – Human coronavírus strain NL63
HCoV-OC43 – Human coronavírus strain OC43
I – Isiaka clan
IZW – Leibniz Institute for Zoo and Wildlife Research
lnL – Maximum Likelihood value
M – Mamba clan
Mb – Membrane protein
MERS – Middle East respiratory syndrome
MHV – Murine hepatitis virus
ML – Maximum-likelihood
mRNA – Messenger ribonucleic acid
NP – National Park
NTPase – Nucleoside-triphosphatase
NTPs – Nucleoside triphosphates
ORF – Open reading frame
P – Pool clan
PCR – Polymerase chain reaction
NS – Non-structural protein
N – Nucleoapsid protein
pre-mRNA – precursor of messenger ribonucleic acid
pAPN – Porcine aminopeptidase N
PEDV – Porcine epidemic diarrhoea virus
RdRp – protease-RNA-dependent RNA polymerase
RNA – Ribonucleic acid
RT-PCR – reverse transcription polymerase chain reaction
SaV – Sapovirus
S – Spike protein
SARS – Severe acute respiratory syndrome
TGEV – Transmissible gastroenteritis virus
VPg – Viral protein genome

## I. INTERNSHIP ACTIVITIES

This master thesis resulted from an internship at the Leibniz Institute for Zoo and Wildlife Research (IZW), in Berlin, starting from the 26th October 2016 until 31st of March 2017 under the supervision of Dr. Marion L. East and co-supervision of Dr. João Nestor das Chagas e Silva.

Throughout the internship in the IZW's Evolutionary Ecology Department I had the opportunity to learn molecular biology techniques that allowed me to work autonomously in the laboratory and execute the work presented in this thesis. During these six months I handled tissue and faecal samples from different wild carnivores and performed the extraction of ribonucleic acid from them. Within the laboratory work I designed primers and performed reverse transcription polymerase chain reactions and polymerase chain reactions. Besides those proceedings, I learned to sequence deoxyribonucleic acid and performed molecular cloning. Following the laboratory work, in order to analyse the results obtained I learned how to work with software like Molecular Evolutionary Genetics Analysis. All the procedures executed during the internship that contributed for the present work are described in the method section of this thesis.

Apart from the work performed for this thesis, my internship at IZW gave me the opportunity to know the research that currently is being done by the institute on wild animals locally, i.e. in Germany, and worldwide in different fields (e.g. parasitology, behaviour, reproduction), and also by researchers from other universities and institutes. It also allowed me to better understand the challenges in research, namely when the subject of study are wild animals in their natural habitat.

In the end, this internship gave me a new and improved vision on the importance of research on wild animals, the restrictions of this work and the importance of sharing the knowledge in order to science to move forward.

## II. INTRODUCTION

In the past years wildlife has been studied extensively throughout several works that focused on a wide range of research topics. These studies not only contribute to increase knowledge on wild fauna and its interaction with the environment, but also help to identify and raise awareness on the threats to the life of these animals with the purpose of preventing drastic population decrease and species' extinction. The decline in population size is known to have a direct impact on the genetic diversity of a given species. A reduction in the pool of genetic information in a species or a population can reduce the scope for evolutionary adjustments and hence survival in a rapidly changing global environment (Ehrlich, 2014; Ceballos, Ehrlich & Dirzo, 2017). A reduction of species diversity has been shown to result in detrimental impacts on both animal and plant communities. The latter can happen, for example, due to the decline of animal populations responsible for plant pollination or dispersal or the reduction of herbivore pressure (Camargo-Sanabria, Mendoza, Guevara, Martinez-Ramos & Dirzo, 2014). Both situations lead to a change in plant communities with loss of species diversity. Thus, the loss of animal populations or species leads to disruptions of ecological networks, affecting the ecosystems on which local and regional human communities depend (Cardinale et al., 2012; Brosi & Briggs, 2013; Ceballos et al., 2017).

Recently, Ceballos *et al.* (2017) discussed an alarming viewpoint on the severity of the Earth's sixth mass extinction that is currently ongoing and highlighted the existence of a very short window, of possibly 20 or 30 years, for effective action against this problem. In the last 100 years about 200 species of vertebrates became extinct. These extinctions occurred at a much higher rate than that estimated to have prevailed over the last 2 million years. During that period, for the same number of species to become extinct it would take 10,000 years (Ceballos et al., 2015). The latest Living Planet Index, a measure of the state of the world's biological diversity based on vertebrate species' trend of population, estimates that in a period of 42 years, between 1970 and 2012, the wildlife in the planet decreased by 58% (World Wide Fund for Nature, 2016).

A species can go extinct by any phenomena that can cause mortality rates to exceed reproductive replacement rates over a sustained period and are usually multi-causational. The rate of extinctions has accelerated in the last centuries mainly due to human action. The main driver of extinctions is habitat loss, fragmentation and degradation. However, other drivers of species extinction include overexploitation, invasive species, pollution, climate change and disease which may act independently or synergistically (Sodhi, Brook & Bradshaw, 2009; MacPhee & Greenwood, 2013).

While the impact of some of these factors on wildlife is entirely dependent on human action (e.g. overexploitation) and easily measured, others such as diseases are difficult to quantify. The lack of knowledge about pathogen diversity, their microbiology, wildlife's susceptibility and host-pathogen relationship has been hindering the understanding of the burden of

diseases in the extinction of species (Sodhi et al., 2009). In addition to these doubts there is the possibility of a pathogen to jump from its original host into a novel species resulting in a spillover event (successful infection with limited or inexistent transmission in the new species) or a host shift (successful infection and transmission in the new species that possibly lead to the emergence of infectious disease that may affect humans, wild and domestic animals) (Longdon, Brockhurst, Russell, Welch & Jiggins, 2014). Research on emerging infectious diseases in humans has revealed that these diseases originated mostly from zoonotic pathogens in wildlife (Daszak, Cunningham & Hyatt, 2000; Jones et al., 2008).

In the last decades more research has been done regarding diseases with impact in wildlife and brought to light the importance of pathogens in previous and current declines of species' populations. While the distemper in African lions (*Panthera leo*) (Roelke-Parker et al., 1996) caused a temporary decline in the species' population, the devil facial tumour disease of Tasmanian devils (*Sarcophilus harrisii*) or the chytridiomycosis (*Batrachochytrium dendrobatidis*) in several amphibian species (Skerratt et al., 2007; Schloegel et al., 2009) are examples of diseases responsible for the continuous decline of wild fauna populations (Murchison et al., 2012).

Continuous monitoring and research are therefore necessary considering the constant evolution and adaptations of both pathogens and hosts.

In this master thesis two studies related to the *Sapovirus* (SaVs) diversity in spotted hyenas (*Crocuta crocuta*) from Serengeti National Park (NP) and the molecular diversity of the *Alphacoronavirus* (α-CoV) receptor, aminopeptidase N (APN), in different wild carnivores will be presented.

# III. SAPOVIRUS IN SPOTTED HYENAS FROM SERENGETI NATIONAL PARK

## 1. LITERATURE REVIEW

### 1.1. *Sapovirus*

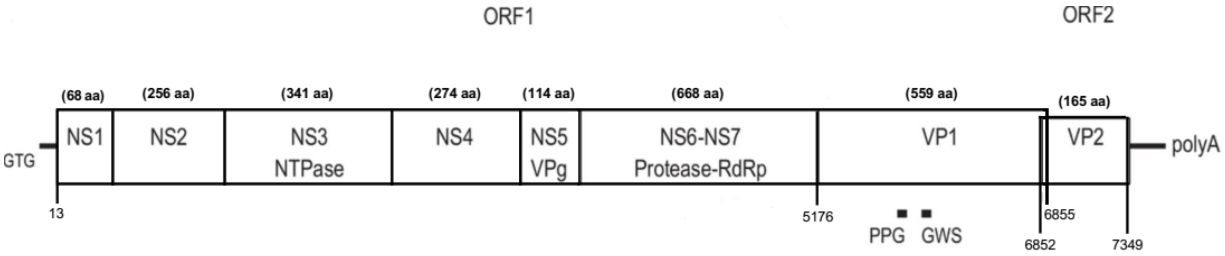
The genus *Sapovirus* is a group of enteric viruses within the family *Caliciviridae* (Mayo, 2002).

SaVs have non-enveloped virions with a linear, positive-sense, single-stranded ribonucleic acid (RNA) genome with approximately 7.1 to 7.7 kb in size including a poly-A tail at the 3'-end. The genome contains two to three open reading frames (ORFs) encoding the viral non-structural proteins (NSs) and the structural proteins (figure 1) (Oka, Wang, Katayama & Saif, 2015).

ORF1 encodes a large polyprotein, which is processed into at least six NSs (NS1, NS2, NS3, NS4, NS5, NS6-NS7) and the major capsid protein, VP1 (Chang et al., 2005). Based on the human SaV Manchester strain (accession number on GenBank: X86560) this ORF is constituted by 6841 nucleotides (from nucleotide 13 to 6852) and the polyprotein has 2280 amino acids (aa) (figure 1). NS5 (114 aa) is a genome-linked protein (VPg) which covalently links to the 5' end of the viral genome playing a critical role for the infectivity and translation of viral RNA (Hosmillo, Chaudhry, Kim, Goodfellow & Cho, 2014). NS6-NS7 (668 aa) is assumed to exist as one single protein with simultaneous proteolytic (NS6) and polymerase (NS7) functions since *in vitro* studies did not show their cleavage in distinct proteins by the viral protease (Oka et al., 2005; Oka, et al., 2006; Fullerton et al., 2007; Robel et al., 2008). Thus, this NS is also known as protease-RNA-dependent RNA polymerase (RdRp). The biological function of SaV's remaining NSs has not been experimentally determined. However, NS3 (341 aa) is presumed to hydrolyse nucleoside triphosphates (NTPs) in order to obtain energy for the virus replication process, since it presents the typical calicivirus nucleoside-triphosphatase (NTPase) amino acid motif (GAPGIGKT) (Pfister & Wimmer, 2001; Oka et al., 2005). VP1 (559 aa) is the major structural protein of this virus, with approximately 60 kDa, being present in the capsid. This encoding region is the most diverse of the SaV's genome and is responsible for the virus antigenicity (Terashima et al., 1983; Katayama et al., 2004; Hansman et al., 2005c; Oka et al., 2015).

ORF1 is followed by ORF2, which is thought to encode the minor structural protein, VP2 (Chang et al., 2005). Based on the Manchester strain of human SaV, this ORF is composed by 497 nucleotides and the VP2 by 165 aa (from nucleotide 6852 to 7349) (figure 1). Some human (Numata, Hardy, Nakata, Chiba & Estes, 1997) and bat (Tse et al., 2012) SaV strains encode an ORF3, however its function is still unknown.

Figure 1. *Sapovirus* genomic organization (Adapted from Oka et al., 2016; Oka et al., 2016)




The diagram is based on the human Sapovirus Manchester strain (accession number on GenBank: X86560). It presents two open reading frames (ORF1 and ORF2), six non-structural proteins (NS) and two structural proteins (VP1 and VP2). Above the genome are presented the predicted sizes of each viral protein in this strain. The numbers below correspond to the predicted size of the genome (beginning and end of both ORFs and the cleavage between NS6-NS7 and VP1). Below the genome are also present the conserved amino acid motifs of VP1 region (PPG and GWS).

### 1.1.1. *Sapovirus* Classification

Based on the genomic or protein sequence sapoviruses can be divided in genogroups and, within these, into genotypes.

The partial RdRp region (corresponding to nucleotides 4273 to 5177 in Manchester strain) or partial VP1 region (corresponding to nucleotides 5074 to 5876 in Manchester strain) can be used to partially characterize a particular SaV strain (Oka et al., 2015).

For the virus classification the RdRp is a less reliable region to use when compared to VP1 region because it has less variability between strains (Oka et al., 2015). For a more accurate and reliable classification the total VP1 region should be sequenced since this is the genome's most diverse part and correlates to the virus antigenicity. Several studies have shown that virus antigenicity differs between genogroups and, in some cases, between genotypes from the same genogroup (Hansman et al., 2005a; Hansman, Natori, Ushijima, Katayama & Takeda, 2005b; Hansman, Oka, Sakon & Takeda, 2007; Oka, Miyashita, Katayama, Wakita & Takeda, 2009). Because of its characteristics the International Calicivirus Conference Committee proposed that in order to designate new genogroups or genotypes it is necessary to sequence at least the complete protein of VP1 region (Oka et al., 2015).

A recent study based on the complete VP1 aa sequence has proposed 15 genogroups (GI to GXV) including both human (GI, GII, GIV and GV) and animal SaV strains: chimpanzee (*Pan troglodytes troglodytes*) – GI; rat (*Rattus norvegicus)* - GII, GXV; pig (*Sus scrofa domesticus*) – GIII, GV-GXI; California sea lion (*Zalophus californianus*) – GV; mink (*Neovison vison*) – GXII; dog (*Canis lupus familiaris*) – GXIII; bat (order Chiroptera) – GXIV) (Oka et al., 2016).

### 1.1.2. Genetic Diversity

The *Sapovirus* genus contains genetically highly diverse viruses (Oka et al., 2016). Currently, it is known that the degree of viral genetic diversity is determined by virus- and host-dependent processes. The viruses adapt and evolve in the sense of increasing their replication, spread the infection and avoid the immune system of the host. At the same time, the infected host tries to adapt its immune response to clear new virus strains that emerge (Sanjuán & Domingo-Calap, 2016).

According to evolutionary theory, viruses can evolve mainly through mutation, natural selection, genetic drifting, migration or recombination (Moya, Holmes & González-Candelas, 2004). The RNA viruses have several characteristics that make them the most susceptible organisms to genetic modifications.

Most RNA viruses lack proofreading activity in the polymerase protein, with exception of coronaviruses (Smith, Sexton & Denison, 2014), making them highly susceptible to produce viral copies with mutations. Those can be (i) deleterious, (ii) neutral or (iii) advantageous to the virus. (i) Deleterious mutations decrease the virus fitness, being able to affect its replication and consequently reduce its population and lead to extinction. (ii) Neutral mutations do not affect the virus fitness at any level. (iii) Advantageous mutations increase the viral fitness enhancing ability of the virus to replicate and, in some cases of high mutation rates, to evade the host's immune response more efficiently (Schotsaert & García-Sastre, 2014; Coppola et al., 2015). RNA viruses are, in fact, the group of organisms with the highest known mutation rates with approximately one mutation per genome, per replication. This might also be the reason why the genomes of RNA viruses are normally small (average of ~9 kb), since high-mutation rates are theoretically expected to limit genome size (Moya et al., 2004). Besides the lack of proofreading activity, the mutations can occur due to spontaneous nucleic acid damage, host enzymes or the presence of genetic elements within the viral genomes whose specific function is to produce new mutations (Sanjuán et al., 2016). The SaVs, as RNA viruses, are highly susceptible to mutations and both synonymous and non-synonymous nucleotide substitutions have been detected in the VP1-encoding region in human individuals during an outbreak (Iwakiri et al., 2009).

RNA viruses' high replication ability is another factor that contributes for their fast evolution and considerable genetic diversity. From one single infectious particle, these viruses can create an average of 100,000 viral copies in 10 hours, building populations of very large size in the host (Moya et al., 2004). Natural selection, the competition between two viruses of the same population with different fitness due to mutation, is most efficient with large populations like the ones presented by RNA viruses. It works as a controller of the evolutionary dynamics allowing the new mutants with increased fitness to out-compete older and inferior alleles (Elena, Miralles & Moya, 1997; Kutnjak, Elena & Ravnikar, 2017).

Migration (or gene flow), corresponds to the transfer of alleles or genes from one population to another. This process can result in virus increasing its host range between species, populations of the same species, individuals of the same population or within a single individual, since from the site of the virus inoculation they can be transported to several tissues. This last path can lead to the development of virus' intra-host spatial variation (Navas, Martín, Quiroga, Castillo & Carreño, 1998; Moya et al., 2004). The outcome of this process in the fitness of the virus is not well studied but some experiments revealed a positive correlation (Miralles, Moya & Elena, 1999).

Finally, recombination is a process where larger genetic changes can occur compared to those that occur through mutation. Recombination involves the exchange of genetic material between two viruses of different strains. These two viruses have to co-infect the host cell at the same time in order to interact during replication and produce a new virus strain. Usually the recombination involves viruses closely related (e.g. from the same genus or the same family) (Fleischmann, 1996). Its frequent occurrence might be advantageous as it can create virus genotypes with high fitness in a shorter period than by mutation alone. It might also be able to eliminate the deleterious mutations from a population avoiding the decrease in its fitness. The recombination rates in RNA viruses are lower than in other organisms but it may have an important role in their evolution (Posada, Crandall & Holmes, 2002). In SaVs the recombinant (or chimeric) strains correspond to viruses that have an inconsistent group between the NS protein-encoding region, which includes the RdRp region, and the VP1 encoding region. Such intra- and intergenogroup have been reported (Oka et al., 2015).

All the mentioned processes involved in virus diversity may contribute to SaVs' quick adaptation to new environments and the production of new strains *in vivo*, explaining the distribution of genetically distinct strains across species (Iwakiri et al., 2009; Kutnjak et al., 2017).

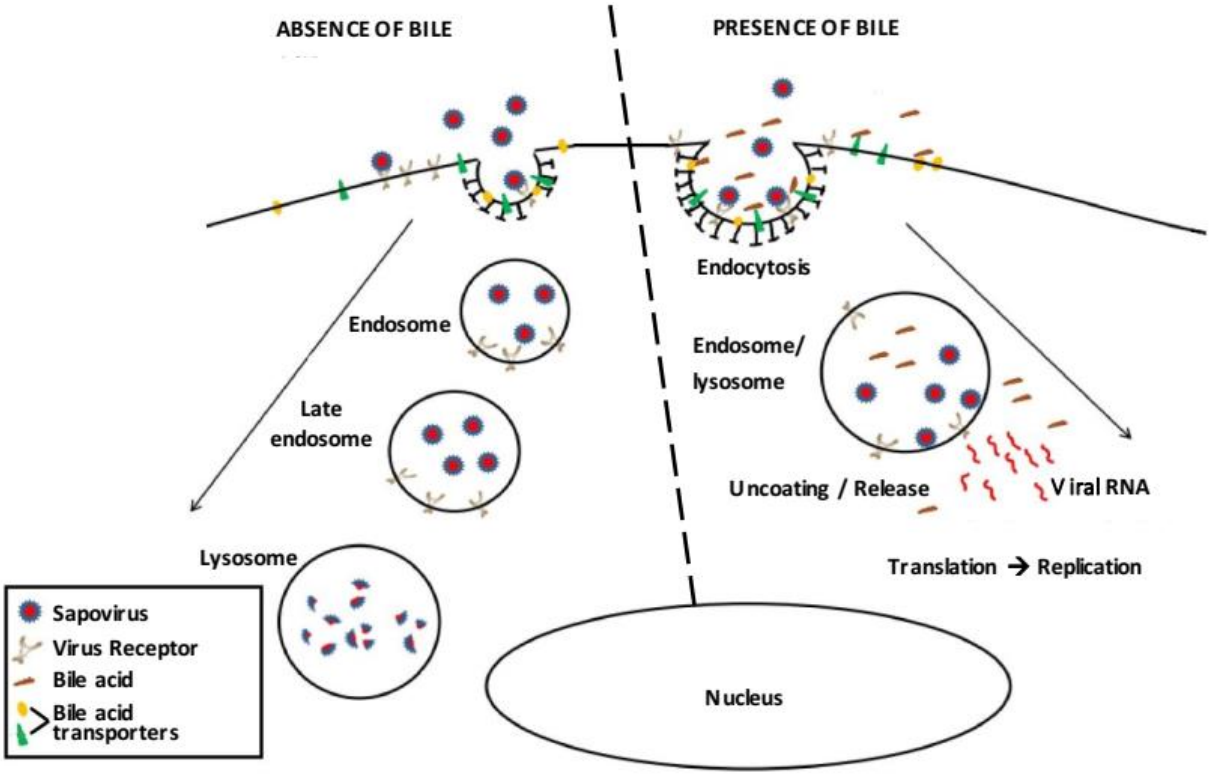
From a different perspective the genetic drift is one of the forces of evolutionary change that has attracted attention to RNA viruses, because of stochastic changes in the allele's frequency in a population. It is of special importance in population bottlenecks, i.e. when a population is drastically reduced and consequently its genetic variation decreases. Genetic drift negatively affects the efficiency of natural selection, as it is more difficult to benefit from the advantageous mutations and eliminate deleterious ones. Thus, the occurrence of genetic drift in a viral population is expected to slow its adaptive evolution (Robertson, 1960; McCrone & Lauring, 2017).

### 1.1.3. Life Cycle

The SaV life cycle has been studied using the porcine SaV strains that can be cultivated such as the Cowden strain (Oka et al., 2015). The primary porcine kidney cells and the continuous porcine kidney cell line are the culture cells in which the virus grows *in vitro* (Parwani, Flynn, Gadfield & Saif, 1991; Farkas, Sestak, Wei & Jiang, 2008).

In order for the replication of SaV to happen the presence of bile acid is required in the culture medium (Chang et al., 2004). *In vivo* the virus replicates primarily in the proximal intestinal tract in the presence of high concentrations of bile acids (Flynn & Saif, 1988). A study from Shivanna, Kim & Chang (2014) revealed the importance of these on the entry of SaV into the host cells. According to this study, the bile acids play a crucial role in the early stages of the virus replication cycle, prior to the release of viral genome into the cytoplasm, as they are thought to have a role in the endosomal escape of SaV. Based on these findings a model for bile acid-mediated porcine SaV replication was proposed (figure 2) where the virus enters the host cell through endocytosis, travels to the late endosomes and, in the presence of bile acids, the virus or its genome is able to escape to initiate the replication. In the absence of bile acids, or when the endosomal pH is not low enough to allow their function, the virus is incapable of escaping the late endosome and is degraded when this fuses with the lysosome.

Figure 2. Possible model of bile acid mediated *Sapovirus* replication (adapted from Shivanna et al., 2014)



As in all positive sense RNA viruses the translation of SaV starts immediately after virus entry into the host cell. The viral genome acts as a messenger RNA (mRNA) template and its VPg interacts with the 4F subunit of the eukaryotic translation initiation factor complex facilitating the viral proteins synthesis. However, not much is known about this process (Hosmillo et al., 2014; Hinnebusch, 2014).

Currently the binding and entry mechanisms of the virus to the host cell, the mechanism and sites of replication and translation of viral RNA remain unknown (Oka et al., 2015).

### 1.1.4. Pathophysiology

The knowledge of SaVs' pathophysiological effect on the host is limited but is assumed that they act in a similar way as rotavirus and norovirus (Model & Burnweit, 2016).

It is known that they are enteric viruses which can cause acute gastroenteritis in humans and animals as sporadic cases or outbreaks of infection in populations (Madeley & Cosgrove, 1976; Svraka et al., 2010; Oka et al., 2015). The virus transmission occurs usually by faecal-oral route, through contact with infected faeces or vomit or with contaminated surfaces, material, food or drinking water, resulting in infection (Hedlund, Rubilar-Abreu & Svensson, 2000; Hansman, Oka, Sakon & Takeda, 2007).

After the incubation period, which in humans ranges from less than a day to 4 days (Lee et al., 2013), the outcome of infection is either a self-limiting gastroenteritis or asymptomatic (subclinical SaV infection) (Matson, Estes, Tanaka, Bartlet & Pickering, 1990; Kirkwood, Clark, Bogdanovic-Sakran & Bishop, 2005; Monica et al., 2007; Gutiérrez-Escolano et al., 2010). SaV infection rarely results in mortality (Lee et al., 2012). Virus shedding in faeces occurs in both symptomatic and asymptomatic individuals and may last 1 to 4 weeks in humans with the possibility of shedding continuing after the symptoms cease (Chiba et al.,1980; Tse et al., 2012). The virus infectious dose and the pathological changes in the infected individual are currently unknown.

Clinical signs include watery, non-bloody diarrhoea associated with vomiting, abdominal pain and/or nausea and are predominantly observed in infants and young children (Nakata et al., 1998; Nakagomi, Nakagomi & Cunliffe, 2013). In domestic pigs the infection causes enteritis in the weaning and post-weaning phases in juvenile animals (Bank-Wolf, Konig & Thiel, 2010).

It is known that an individual previously infected with SaV that has serum antibodies against that strain will have a reduced frequency of SaV infection and illness by that strain and all the antigenically homologous strains (Nakata, Chiba, Terashima, Yokoyama & Nakao, 1985). However, a symptomatic reinfection can occur if a host is infected with a SaV strain that belongs to a different genogroup/genotype (Harada et al., 2012). Besides this, the host's immunological response and protective immunity to SaV remain unclear (Oka et al., 2015).

### 1.1.5. Virus Detection

Diagnosis of SaV infection requires the virus detection which can be achieved through different methods.

Using electron microscopy is possible to identify virus' particle by the characteristic "Star of David" surface morphology of viruses in the *Caliciviridae* family, which distinguishes SaVs from other gastrointestinal virus (e.g. rotavirus, astrovirus, adenovirus) (Madeley, 1979).

Enzyme-linked immunosorbent assays have been used for SaV's antigen detection in human samples but, due to the difficulty in detecting antigenically diverse SaV strains and current lack of commercial availability, this method is not widely used for diagnosis in humans and other species (Nakata, Estesl & Chiba, 2005; Hansman et al., 2006).

The most sensitive diagnostic method is the detection of SaV by reverse transcription polymerase chain reaction (RT-PCR). The primers used in the diagnostic context in humans aim to amplify the partial RdRp (nucleotides 4273 to 5177 in Manchester strain), RdRp-VP1 junction (nucleotides 5074 to 5261 in Manchester strain) or partial VP1 region (nucleotides 5074 to 5876 in Manchester strain). When amplifying the partial RdRp region using primers which target its conserved motifs, other gastrointestinal viruses besides SaVs may also be amplified in the process (Jiang et al., 1999; Ludert, Alcalá & Liprandi, 2004). The region with highest detection rate, and therefore the first choice for SaV screening from clinical specimens, is the RdRp-VP1 junction (Vinjé et al., 2000; Harada et al., 2009). However, the VP1 region is the ideal to amplify when the purpose is to sequence the products and to genotype the strain responsible for the infection (Kitajima et al., 2010). Thanks to its specificity and sensitivity real time RT-PCR is considered a routine method for detection of viruses from family *Caliciviridae* from human clinical specimens (faeces). However, it is unclear whether the primers used in this process are able to detect the human SaV strains from all genogroups

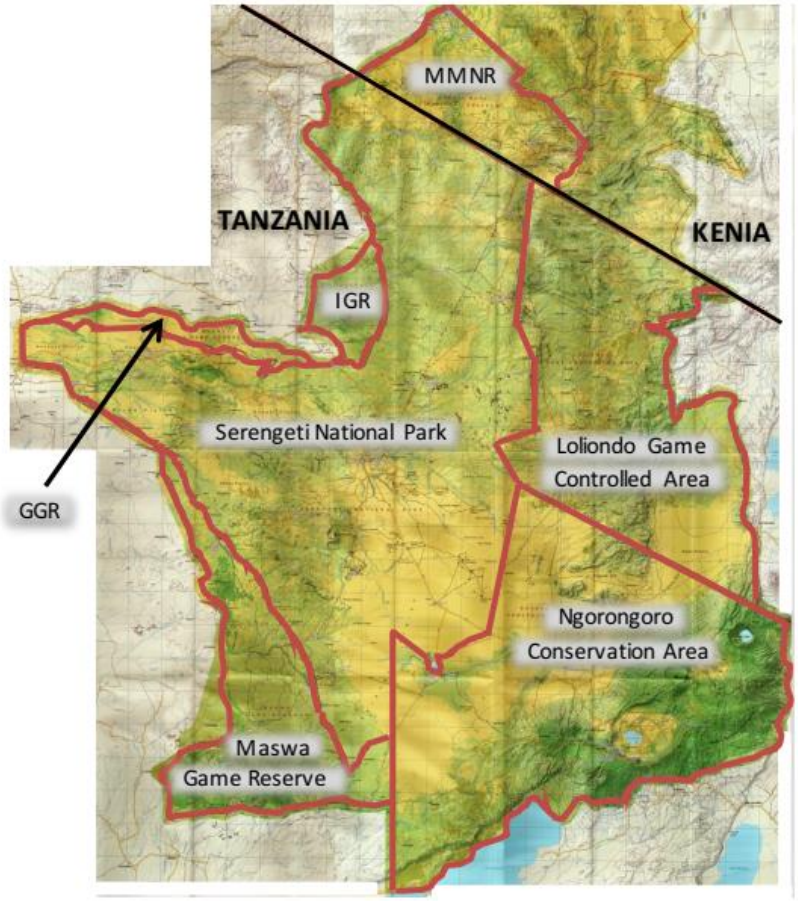
Next generation sequencing techniques have also been used to detect SaV (e.g. faeces from dogs or California sea lions) (Li et al., 2011a; Li et al., 2011b) but they are not used for routine clinical diagnosis due to their high cost (Oka et al., 2015).

SaV particles were detected for the first time in human diarrheic stool samples in 1976 in the United Kingdom (Madeley & Cosgrove, 1976) and since then have been detected worldwide in both humans and animals such as the pig (Saif, Bohl, Theil, Cross & House, 1980; Guo et al., 1999), mink (Guo, Evermann & Saif, 2001), dog (Li et al., 2011a), California sea lion (Li et al., 2011b), bat (Tse et al., 2012), chimpanzee (Mombo et al., 2014) and rat (Firth et al., 2014). Most recently this virus was detected for the first time in wild carnivores species in Serengeti NP, namely bat-eared fox (*Otocyon megalotis*), African lion (*Panthera leo*) and spotted hyena (*Crocuta crocuta*) (Olarte-Castillo et al., 2016).

### 1.2. Serengeti National Park

The Serengeti NP, Loliondo Game Controlled Area, Ngorongoro Conservation Area, Maswa Game Reserve, Grumeti Game Reserve and Ikongoro Game Reserve (in Tanzania) and the Maasai Mara National Reserve in Kenya cover the great majority of Serengeti Mara Ecosystem (figure 3) (Sinclair & Arcese, 1995). This ecosystem is located in part of northern Tanzania and southwestern Kenya and is considered one of the most important conservation areas of the world because of its wildlife diversity and abundance (Boone, Thirgood & Hopcraft, 2006). At the same time it is a human multi-ethnic area with over 30 tribes living there along with their domestic animals (cattle, goats, pigs, donkeys and poultry) (Campbell & Hofer, 1995; Loibooki, Hofer, Campbell & East, 2002).

Figure 3. Map of Serengeti Mara Ecosystem and its protected areas (adapted from UNESCO, 2010)




Maasai Mara National Reserve (MMNR); Grumeti Game Reserve (GGR); Ikongoro Game Reserve (IGR). The red line limits the protected areas. The black line limits the geographic border between Kenia and Tanzania.

The seasons and landscape of this region are defined by seasonal rainfall which leads to water and vegetation variations within the Serengeti Mara Ecosystem. As consequence an annual migration occurs involving wildebeests (*Connochaetes taurinus*) zebras (*Equus burchelli*) and Thomson's gazelles (*Gazella thomsoni*) (Estes, 1976; Hofer & East, 1993a). Besides the migratory species, the protected areas house resident herbivores such as elephants (*Loxodonta africana*), hippopotamus (*Hippopotamus amphibius*), elands (*Tragelaphus oryx*), impalas (*Aepyceros melampus*), giraffes (*Giraffa camaleopardis*), warthogs (*Phacochaerus aethiopicus*), among others.

The avifauna in Serengeti is also very rich with estimated 500 species, some of restricted range like Fisher´s lovebird (*Agapornis fischeri*), Karamoja apalis (*Apalis karamojae*) and grey-crested helmet shrike (*Prionops poliolophus*).

The highest diversity and biggest populations of wild carnivores in African savanna are also found in this region with 26 species described (table 1) (Sinclair & Arcese, 1995, Stattersfield, Crosby, Long & Wege, 1998). In the Serengeti NP, the major protected area that lies in its centre of the Serengeti Mara Ecosystem, in Tanzania, the most common large carnivore is the spotted hyena (Hofer & East, 1993a).

Table 1. Wild carnivore species that inhabit the Serengeti Ecosystem (Adapted from Frame, 1986; Serengeti National Park's Official Site, 2000; IUCN, 2017)

| **Classe** Mammalia **Order** Carnivora | | |
|---|---|---|
| **Suborder** | **Family** | **Species** |
| Caniformia | Canidae | African wild dog (*Lycaon pictus*)<br>Bat-eared fox (*Otocyon megalotis)*<br>Black-backed or Silver-backed jackal (*Canis mesomelas*)<br>Golden or Common jackal (*Canis aureus*)<br>Side-striped jackal (*Canis adustus*) |
| | Mustelidae | Ratel or Honey badger (*Mellivora capensis*)<br>Zorilla or Striped polecat (*Ictonyx striatus*) |
| Feliformia | Felidae | Caracal (*Felis caracal*)<br>Cheetah (*Acinonyx jubatus*)<br>Leopard (*Panthera pardus*)<br>Afican lion (*Panthera leo*)<br>Serval (*Leptailurus serval*)<br>African wild cat (*Felis lybica*) |
| | Hyaenidae | Aardwolf (*Proteles cristatus*)<br>Spotted hyena (*Crocuta crocuta*)<br>Striped hyena (*Hyaena hyaena*) |
| | Viverridae | African civet (*Civettictis civetta)*<br>Large-spotted genet (*Genetta tigrina*)<br>Small-spotted genet (*Genetta genetta*) |
| | Herpestidae | Banded mongoose (*Mungus mungo*)<br>Marsh mongoose (*Atilax paludinosus*)<br>White-tailed mongoose (*Ichneumia albicauda*)<br>Dwarf mongoose (*Helogale parvula*)<br>Ichneumon or Egyptian mongoose (*Herpestes ichneumon*)<br>Slender or Black-tipped mongoose (*Herpestes sanguinea*) |

### 1.2.1. Spotted Hyenas (*Crocuta crocuta*) of Serengeti National Park

Spotted hyenas (suborder Feliformia, family Hyaenidae) are spotted coat and round ears wild carnivores with little sexual dimorphism since both males and females present a prominent structure as external genitalia. While in males that structure corresponds to the penis, in females this structure is the clitoris from which they urinate, copulate and give birth. Both structures are similar in form and size and are presented erected as signal of submission during greetings between spotted hyenas (East, Hofer & Wickler, 1993; Hofer & East, 2013).

This species is classified as "Least Concerned", the lowest conservation risk, in the Red List of Threatened Species from the International Union for Conservation of Nature due to its wide distribution and population's numbers above 10,000 mature individuals, but their

population is declining due to habitat loss and persecution. They inhabit an extensive area of sub-Saharan Africa, especially West and Central Africa, and its estimated global population ranges between 27,000 and 47,000 being the largest known populations placed in the Serengeti ecosystem, in Tanzania and Kenya, followed by Kruger National Park in South Africa (Bohm & Höner, 2015).

The spotted hyena population in Serengeti NP has been study subject for many years (Kruuk, 1972; Hofer & East, 1993a,b,c; Olarte-Castillo et al., 2016). These animals live in large stable clans constituted by an average of 47 members, including adults and subadults. Females remain in the clan in which they are born whereas males typically disperse. Within a clan, females are social dominant over immigrant males (East & Hofer, 2001). Clan size is not limited by resources on the territory due to their foraging behaviour (Hofer & East, 1993a).

The females usually have one or two cubs per litter that remain in the communal den of the clan's territory around 12 months. During this time they are dependent on their mother's milk (Hofer, Benhaiem, Golla & East, 2016). At the age of approximately 3 years, males leave their natal clan's territory. The majority of reproductively active males in a clan are immigrants (Kruuk, 1972).

#### 1.2.1.1. Feeding Behaviour

Spotted hyena's main preys are the migratory ungulates of Serengeti Mara Ecosystem: wildebeests, zebras and Thomson's gazelles. The migratory movements of these animals produce large fluctuations in local prey abundance within clan territories throughout the year forcing spotted hyenas to regularly undertake short-term, long distance foraging trips from their clan's territory in order to feed in areas containing high densities of migratory herbivores. This foraging mode has been termed commuting and commuting trips are thus regular patterns of movement between a central point and foraging sites outside the clan's territory during periods when migratory prey are not present in high numbers in a clan's territory.

When their mothers are absent on commuting trip cubs are usually left at the clan's communal den, as they are able to survive several days without being nursed. With approximately 12 months of age cubs may accompany their mothers on commuting trips (Hofer & East, 2013).

#### 1.2.1.2. Territory

The clan territory serves as breeding and feeding area and contains a communal den that functions as the social centre of the clan and the place where females rear their cubs (Kruuk, 1972; Mills, 1990).

The underground burrows of a communal den are only accessible to cubs and are essential for their survival, providing cool daytime resting and also reducing the risk of predation (Kruuk, 1972; Hofer & East, 1993b). The communal den is situated inside the clan territory but its location can change frequently within its area (Hofer et al., 1993a). Long distance relocation of communal dens, outside their territory, is possible but rare probably due to the high risk of losing cubs during the transfer through exposing to potential predation by lions. Thus, spotted hyenas have a unique characteristic among terrestrial carnivores as they simultaneously maintain and defend their territory and travel long distances to foraging sites.

Spotted hyenas' territory boundaries are defended by its clan members through scent marking including defaecations at latrines and the deposition of anal gland scent on vegetation, vocal displays ('whooping'), territorial patrols, aggressive expulsions of non-residents, and clashes with neighbouring clans.

Throughout the year the clans' territories are crossed by non-resident spotted hyenas which can be classified as (i) commuters in transit, (ii) commuters at kills and (iii) intruders from neighbouring clans. (i) The commuters in transit are the animals who cross the territory using recognized paths and do not utilize the resources within it. These commuters are tolerated by the residents. (ii) The commuters at kills enter the territory to forage and consume food that could be used by residents. The intrusion pressure from these animals increases when migratory herds are present in the territory (East & Hofer, 1991). Commuters are usually not attacked by residents while moving through a territory and are only attacked if they do not relinquish kills to territory owner. In order to minimize feeding competition with residents, commuters forage at a distance of more than 3 kilometres from the resident clan's communal den, since spotted hyenas are able to detect kills over distances of 3 to 4 kilometres. (iii) Intruders from the neighbouring clans are the ones that enter the territory to challenge ownership of a kill and/or part of a territory and its resources. In response territorial clashes are initiated which can rarely culminate in fatalities (Hofer & East, 2013).

### 1.2.2. Serengeti National Park as a Research Subject

For many years the Serengeti National Park has been studied because of its great biodiversity. Since 1987 the Spotted Hyena Research Project, which is now based at the department of Evolutionary Ecology of Leibniz Institute for Zoo and Wildlife Research (IZW) has conducted a long-term research program based in the centre of Serengeti NP. The project has published several behavioural and epidemiological studies, which have reported the detection of pathogens such as coronavirus (CoV) (Goller, Fickel, Hofer, Beier & East,

2013; East et al., 2004), kobuvirus (Olarte-Castillo et al., 2015), SaV (Olarte-Castillo et al., 2016) or canine distemper virus (Nikolin et al., 2016) in wild carnivores from that region. A special emphasis is given to spotted hyenas in this project, since they are the most common large carnivore living in the Tanzanian part of the Serengeti Mara Ecosystem. In this context several hundred spotted hyenas from three major clans ("Isiaka" [I], "Mamba" [M] and "Pool" [P]) have been closely monitored during their life span and individually identified by their spot patterns, scars and other particular features like natural ear-notches (Hofer & East, 1993a). They are also sexed based on body outline differences, reproductive status (e.g. lactation), hair and histological samples examination and, from the third month onwards, using the dimorphic shape of the phallic gland (Frank et al., 1990).

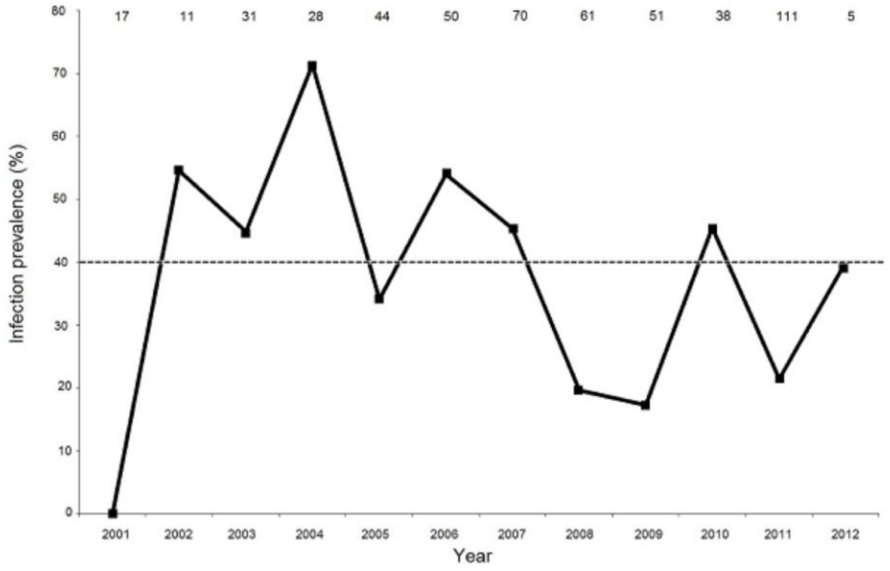
#### 1.2.2.1. *Sapovirus* in Serengeti National Park

In 2016, a paper published by Olarte-Castillo *et al*. presented the results of a long-term study on SaV in wild carnivores in the Serengeti ecosystem. The study took place in Serengeti NP between February 2001 and March 2012 and resulted in the first report of SaV in spotted hyenas, African lions and bat-eared foxes.

An extensive work on spotted hyenas faecal and tissue samples, from the three different clans (I, M and P), was performed and revealed SaV infection prevalence of 34,8% during the study period. Throughout that time three outbreaks of SaV infection occurred in years 2003/2004, 2006/2007 and 2010, corresponding to the years when the infection prevalence was equal or above 40% and sampling size was of at least 20 individuals (figure 4). The first outbreak from 2003/2004 revealed the highest infection prevalence (above 72,4%) in comparison to the subsequent ones which had considerably lower infection prevalence. All the outbreaks were followed by a decrease on SaV's prevalence suggesting an increasing immunity in the previously affected populations.

It was hypothesized that during outbreaks the herd immunity against the genetic strain circulating at the time in the study population increases. Herd immunity, a population's resistance to the spread of a disease when a large proportion of its members become immune after infection, probably has happened in 2003/2004 when a high prevalence of infection was registered. The large number of individuals that were infected during this outbreak seroconverted and hence became immune to the antigenically homologous strains. Because they were more numerous than the number of susceptible individuals in the population, the likelihood of a spotted hyena that was shedding the virus passing on the disease to susceptible animals decreased.

Figure 4. Prevalence of *Sapovirus* infection in spotted hyenas from Serengeti National Park between 2001 and 2012 (Olarte-Castillo et al., 2016)




The numbers above the plot correspond to the size of the sample for each year.

In order to herd immunity to establish, infection must induce solid immunity which will not be the case if an antigenically different strain emerges (Gordis, 2014). As previously mentioned the transmission happens normally via faecal-oral route and not much is known about immunity response to SaV infection. The identification of SaV strains from chimpanzee (Mombo et al., 2014), pig (Martella et al., 2008; L'Homme, Brassard, Ouardani & Gagné, 2010; Scheuer et al., 2013) and rodents (Firth et al., 2014) that phylogenetically clustered closely to human strains in GI, GV and GII, respectively, suggests the possibility of interspecies transmission of the virus. However, so far the cross-species transmission and the existence of host reservoirs have not been demonstrated. According to the genetic analysis of SaV partial RdRp region in strains from spotted hyenas (Olarte-Castillo et al. 2016) the viruses found in Serengeti NP formed a separate group from the SaVs worldwide. Besides that, the viruses from spotted hyenas formed a monophyletic group separated from the others of African lions and bat-eared fox, suggesting that there is no direct transmission of SaVs between the different species in Serengeti NP.

Following this hypothesis the outbreaks registered in 2006/2007 and 2010 could have resulted from infection from the same virus strain that was present in the previous outbreak had the level of herd immunity decreased, for example through the recruitment of young immunologically naïve animals into the population or as a result of a decline antibody titres among exposed animals over time (or both processes). Alternatively, these later outbreaks resulted of the appearance of one or more genetically different SaV strains that evade existing antibodies.

### 1.3. Aim of the Study

The aim of the present study builds on the results obtained by Olarte-Castillo *et al*. (2016), namely to answer the question as to whether the outbreaks of infection detected in the spotted hyena population resulted from a waning of herd immunity or the emergence of new strains that were antigenically different from those in earlier outbreaks.

When in place, herd immunity limits or prevents the virus' circulation within a population leading to its limited and inefficient transmission or even a dead end infection (Longdon et al., 2014). The high ability of RNA viruses to genetically change through different mechanisms (mutation, natural selection, genetic drift, migration or recombination) may lead to the development of new virus strains even in these extreme cases. The resulting new strains, in some cases, may be antigenically different from the original which gives them an evolutionary advantage as they are able to avoid the immune system (Moya et al., 2004). As consequence the new strain can spread the infection again, by infecting the originally susceptible individuals and re-infecting the individuals previously considered as immune post-infection, causing an outbreak of disease.

Olarte-Castillo et al. (2016) studied the genetic diversity of SaV in Serengeti NP by sequencing the partial RdRp encoding region, but this gene fragment doesn't provide relevant information for the determination of the infecting strains antigenic type.

In the present work, in order to test the hypothesis that the outbreaks that followed the one from 2003/2004 resulted from the emergence of antigenically different SaV strains, I focused on sequencing the SaV's VP1 encoding region, responsible for the virus antigenicity, using conventional RT-PCR methods and by applying a phylogenetic analysis to examine the genetic relationship of different homologs sequences and their hypothetical ancestors. The samples included in the study were those known to have adequate virus RNA for the analysis.

## 2. MATERIAL AND METHODS

### 2.1. Samples

For the present work an initial sampling of three faecal samples from distinct known spotted hyenas (I 579, I 600 and M 679) were used to extract and sequence SaV RNA. They were collected in Serengeti NP, Tanzania, shortly after deposition and frozen at -80ºC for storage and transportation after being thoroughly mixed and divided in aliquots. Both I 579 and I 600 were faecal samples from spotted hyenas belonging to clan “Isiaka”, while sample M 679 was from a spotted hyena in clan “Mamba”. Faecal samples I 579 and M 679 were collected in January 2011 and I600 was collected in July of the same year.

All three individuals were previously shown to be infected by SaV through these samples in a study on SaV infection across years and SaV genetic diversity. In that study partial RdRp region of the SaVs that infected I 600 and M 679 was sequenced. The sequences obtained for I 600 (700 nucleotides) and for M 679 (208 nucleotides), accession numbers on GenBank: KT777559 and KT777560, respectively, are a complete match in M 679’s full length (Olarte-Castillo et al., 2016).

### 2.2. RNA Extraction

RNA was extracted from 200µL of faecal suspension in diethylpyrocarbonated (DEPC) treated water using the QIAmp MinElute Virus Spin Kit (QUIAGEN, Hilden, Germany), according to manufacturer’s instructions. The product of extraction was stored at -80ºC between assays to guarantee the stable preservation of the RNA.

### 2.3. RT-PCR, Electrophoresis and Sequencing

In order to amplify, detect and sequence the SaV genome, RT-PCRs were performed followed by agarose gel electrophoresis and sequencing whenever bands corresponding to amplicons of the expected size were observed.

During RT-PCR, the virus genome is converted into its complementary deoxyribonucleic acid (cDNA), using a reverse transcriptase, which is then amplified by standard polymerase chain reaction (PCR). This can be executed as a two-step or a one-step process. In the two-step RT-PCR, for the reverse transcription can be used non-specific or specific primers and the resulting cDNA is added to a second tube where the PCR is performed with gene-specific primers. In the one-step RT-PCR both processes use the gene-specific primers and take place in the same tube (Farrell, 2010).

In this study all the RT-PCRs were executed as a one-step process using the One *Taq*® One Step RT-PCR kit (New England Biolabs, Ipswich, Massachusetts, USA) adjusting the user manual’s instructions to a total reaction volume of 12,5 µL. Therefore, each RT-PCR required 6,25 µL of reaction mix, containing the deoxynucleotide triphosphates (dNTPs); 3,75 µL of RNAse free water; 0,5 µL of the enzyme mix, which includes the reverse transcriptase and

the DNA polymerase. To this were added 0,5 µL of each primer (10µM) and 1 µL of the RNA previously extracted. The reaction mixture was placed in the thermocycler (peqSTAR, Peqlab, Erlangen, Germany) and the one step RT-PCR was performed following the instructions on table 2.

Table 2. Thermocycling conditions for RT-PCR (Adapted from New England Biolabs, Inc., 2016)

| Cycle step | Temperature | Time | Cycles |
|---|---|---|---|
| Reverse transcription | 48ºC | 30 minutes | 1 |
| Initial denaturation | 94ºC | 1 minute | 1 |
| Denaturation | 94ºC | 15 seconds | 35 - 40 |
| Annealing | * | 30 seconds | |
| Extension | 68ºC | 1 min/kb | |
| Final extension | 68ºC | 5 minutes | 1 |
| hold | 8ºC | ∞ | 1 |

* Annealing temperature registered for each pair of primers used for SaV amplification in annex 1

Finished the RT-PCR, 5 µL of the resulting mixture were add to bromophenol Blue-based loading dye and placed into the well of a 1,5% agarose gel. The DNA fragments resulting from RT-PCR were then separated by their sizes using agarose gel electrophoresis, at 100V for 35 minutes, and the results visualized under ultraviolet light.

After confirmation of SaV's presence in the sample using the primers Cali2F and Cali2R that target the RdRp region, other primers (table 3 and annex 1) were used to obtain a fragment for the sequencing of VP1 encoding region.

Whenever a single amplicon of the expected size was detected on agarose gel electrophoresis, the RT-PCR product was submitted to a purification step. In this process the exonuclease I (20U/µL) (Thermo Fisher Scientific, Waltham, Massachusetts, USA) removed the primers and free dNTPs from the mixture, being its function posteriorly inactivated by the FastAP Thermosensitive Alkaline Phosphatase (1U/µL) (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The protocol consisted on adding 5 µL of a clean-up MasteMix (10 µL of exonuclease I; 25 µL of FastAP; 465 µL water) to the RT-PCR product and submitted to the cycle presented on table 4. In the end of this clean-up step the mixture contained only the purified RT-PCR products ready for sequencing.

Table 3. Primers used on *Sapovirus* amplification and sequencing

| Primer | Sequence (5' to 3') | Sense | Position* | Reference |
|---|---|---|---|---|
| Cali2F | CAG TGA CAG CCA CAT CCT TG | Forward | 4388 - 4407 | Olarte-Castillo et al., 2016 |
| Cali2R | AGC ACT GCA GCA GCA AAG TA | Reverse | 4576 - 4595 | Olarte-Castillo et al., 2016 |
| Cali2Rfwd | TAC TTT GCT GCT CA GTG CT | Forward | 4576 - 4595 | Adapted from Olarte-Castillo et al., 2016 |
| SV-F13 | GAY YWG GCY CTC GCY ACC TAC | Forward | 5074 – 5094 | Okada et al., 2002 |
| SV-F14 | GAA CAA GCT GTG GCA TGC TAC | Forward | 5074 – 5094 | Okada et al., 2002 |
| SV-F22 | SMW AWT AGT GTT TGA RAT G | Forward | 5154 – 5172 | Okada et al., 2002 |
| SaV 1245Rfwd | TAG TGT TTG ARA TGG AGG G | Forward | 5159 – 5177 | Sano et al., 2011 |
| SV-G1-R | CCC BGG TGG KAY GAC AGA AG | Reverse | 5561 – 5580 | Okada et al., 2002 |
| SV-R2 | GWG GGR TCA ACM CCW GGT GG | Reverse | 5572 – 5591 | Okada, Shinozaki, Ogawa & Kaiho, 2002 |
| SV-R13 | GGT GAN AYN CCA TTK TCC AT | Reverse | 5857 – 5876 | Okada et al., 2002 |
| SV-R14 | GGT GAG MMY CCA TTC TCC AT | Reverse | 5857 – 5876 | Okada et al., 2002 |
| "New F" | AAT KTG AAC TAT GAY CAK GCW CKC GC | Forward | 5062 - 5087 | Original |

* Position on nucleotide sequence based on Manchester strain (accession number on GenBank: X86560.1)

Table 4. Thermocycling conditions for clean-up prior to sequencing (Thermo Fisher Scientific, 2016)

| Cycle step | Temperature | Time | Cycles |
|---|---|---|---|
| Incubation | 37ºC | 15 minutes | 1 |
| Stop reaction | 85ºC | 15 minutes | 1 |
| hold | 8ºC | ∞ | 1 |

To this step followed the bidirectional sequencing of the RT-PCR products. Firstly, a new PCR was performed using 1-3 µL of the template; 2µL of BrightDye® Terminator 5X Sequencing Buffer; 0,5 µL of BrightDye® Ready-reaction Premix; 0,5 µL of BDX64 Sequencing Enhancement Buffer (NimaGen, Nijmegen, The Netherlands) and 1µL of corresponding primer, with the thermal cycling performed following the instructions on table 5.

Table 5. Thermocycling conditions for sequencing reaction (Adapted from Nimagen, 2015)

| Cycle step | Temperature | Time | Cycles |
|---|---|---|---|
| Initial denaturation | 96ºC | 3 minutes | 1 |
| Denaturation | 96ºC | 10 seconds | 30 |
| Annealing | * | 5 seconds | |
| Extension | 60ºC | 2 minutos | |
| hold | 8ºC | ∞ | 1 |

* Annealing temperature used for each primer corresponds to the temperature in the amplification (annex 1)

In this reaction both dNTPs and dideoxynucleotide triphosphates (ddNTPs) were used. The ddNTPs are modified dNTPs that present a hydrogen group on the 3' carbon where usually is a hydroxyl group (HO). That change stops the synthesis of a sequence by preventing the addition of further nucleotides. In addition, for sequencing purpose, ddNTPs are labelled with different colour fluorescent dyes. As result several DNA sequences with different lengths were synthetized. This process was then followed by a clean-up step using the D-Pure DyeTerminator Removal kit (NimaGen, Nijmegen, The Netherlands) according to manufacturer's instructions, with the purpose to remove all the contaminants (salts and unincorporated fluorescent ddNTPs) from the previous cycle sequencing reaction. Lastly, a capillary electrophoresis was performed in the 3130 Genetic Analizer (Applied Biosystems, Darmstadt, Germany) where the different wavelengths of the four fluorescent dyes were detected by laser and resulted in a chromatogram (Russel, 2010).

### 2.4. Phylogenetic Analysis

The nucleic acid sequences were edited using the Chromas 2.6.4 program (Technelysium Pty Ltd, Australia) and their homology with the SaV was verified using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1997) followed by the phylogenetic analysis on MEGA version 6 (MEGA 6) (Tamura, Stecher, Peterson, Filipski & Kumar, 2013).

For the phylogenetic analysis other SaV' sequences retrieved from GenBank were added. At least one sequence representative of each genogroup (GI-GXV), according to Oka et al. (2016) latest proposal, was included in this analysis. These sequences from different species had the following accession numbers: GI (human - DQ366345), GII (human - AY646855); GIII (pig - KT922087); GIV (human - DQ058829); GV (human - AY646856, California sea lion - JN420370); GVI (pig - AY974192); GVII (pig - AB221130); GVIII (pig - KC309419); GIX (pig - KC309418); GX (pig - AB242873); GXI (pig - DQ359100); GXII (mink - AY144337); GXIII (dog - JN387135); GXIV (bat - JN899075); GXV (rat - KJ950878). All sequences were aligned using the MUSCLE method (Edgar, 2004a,b) based on the sequences' codons. The

presence of typical aa motifs from SaV was verified by translating the nucleotide sequence into a protein sequence on MEGA 6.

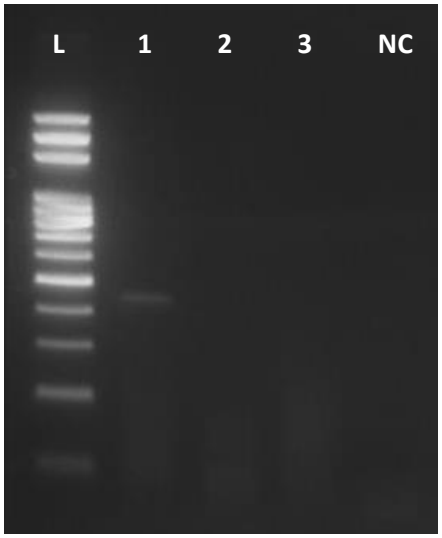
The phylogenetic relationship between the nucleotide sequences of the different strains was estimated using the maximum-likelihood (ML) method based on the partial VP1 encoding region obtained in this work. The best model to construct the phylogenetic tree was determined using the Bayesian information criterion (BIC), Akaike Information Criterion, corrected (AICc) and the Maximum Likelihood value (lnL). The Tamura 3-parametre model (Tamura, 1992) plus gamma distribution plus invariant sites (T92+G+I) was used for that purpose with 1000 bootstrap replicates as statistical support (figure 7). The bootstrap values, indicators of the reliability of the cluster descending from each node, are quoted on the branches. They are presented in percentage and the higher the number, the more reliable is the estimate of the taxa descending from that node. Bootstrap values <70% were not considered reliable.

## 3. RESULTS

The screening RT-PCR performed with primer pair Cali2F/Cali2R confirmed the presence of SaV RNA in the three samples.

The subsequent RT-PCRs performed with the remaining primer pairs (annex 1) did not produce the desired results with exception for primer pair SaV1245Rfwd/SV-R2. These latter primers amplified a fragment of sample I 600, resulting in the presence of a positive band on gel electrophoresis, but the other two samples were negative (figure 5).

Figure 5. Picture of gel electrophoresis' result of SaV amplification with primers SaV1245Rfwd and SV-R2 (Original)




L – 100 bp DNA Ladder; 1 - sample I 600; 2 - sample M 679; 3 – sample I 597; NC – negative control. A band with approximately 430 bp is present in sample I 600.

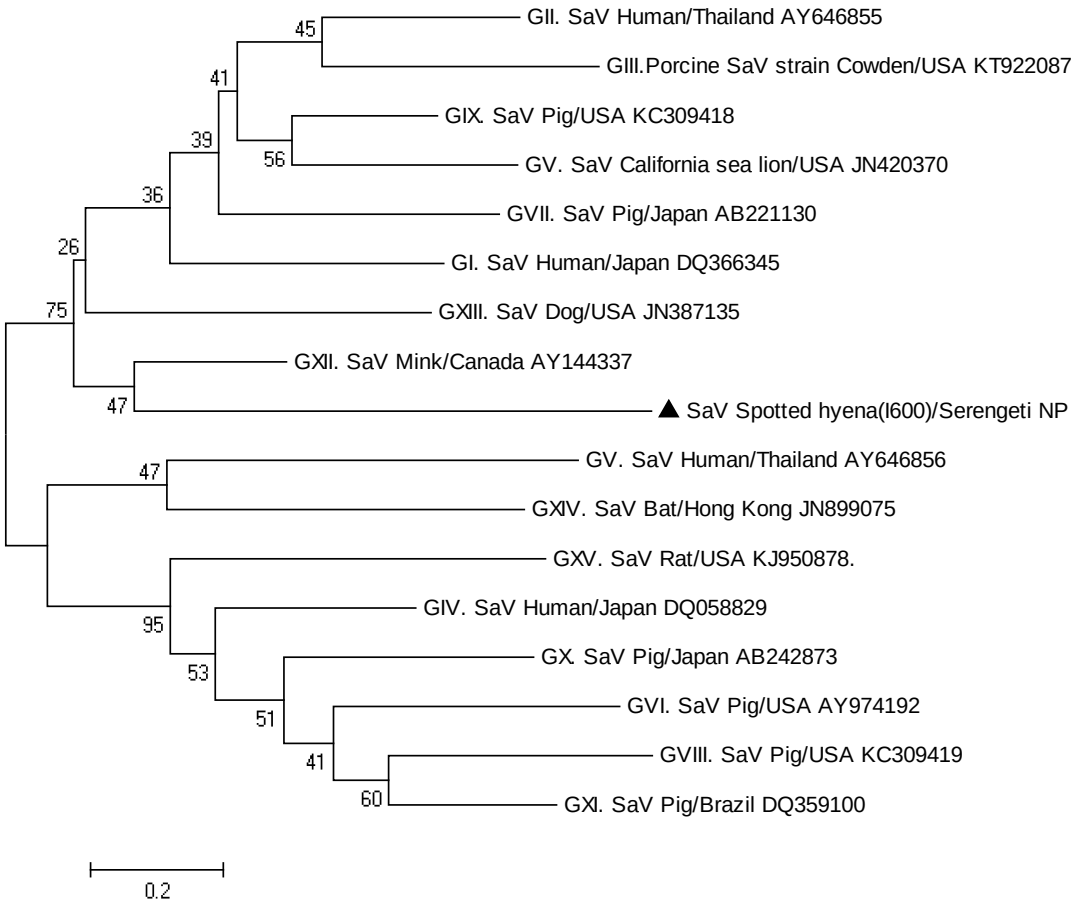
The genomic sequence from I 600 assembled on MEGA6 (Tamura et al., 2013) is constituted by 351 nucleotides (figure 6), corresponding to nucleotides 5242 to 5592 of Manchester strain. The corresponding aa sequence (116 aa) presents one of the typical aa motifs for VP1 (PPG) (figure 1).

Based on the sequence obtained, the unrooted phylogenetic tree on figure 7 was constructed. The tree presented a poor statistical support for the majority of the clusters, however, two separate clusters presented a bootstrap value >70%. At the top of the tree a cluster containing the SaVs' sequences from genogrups I, II (human), III (pig), V (California sea lion), VII, IX(pig), XII (mink), XIII (dog) and the sequence obtained for spotted hyena's SaV was supported by a bootstrap value of 75%. In the bottom of the tree the sequences from genogroups IV (human), VI, VIII, X, XI (pig), XV (rat) were clustered with a statistical support of 95%. An isolated group formed by a human (GV) and bat (GXIV) SaVs was also found in the phylogenetic tree.

Figure 6. Partial sequence of *Sapovirus*' VP1 encoding region found in spotted hyena (I 600) (Original)

CGGGCCCTGCATCTGGGTGATGGGATTGTAGTTAACCCCCAGCCCCCAGT
CCAACAAGCGGGCGCAACAGCACTAGCAGTTGCCACTGGCGCCATTGAGT
CTAACGTTCCGTCTGTATTGCGGAACTTCTTCACACTGTCCCATTCTTGGGT
GTGGTCTTCCCGTGCTGGTCCTGGTTCAATGATTGGTTCGATGCGGCTAG
GCCCCGGCAACAATCCCTATGTTAGCCATTTGTCAGCAATGTATGCTGCGT
GGGGCGGTAGCATGGACGTGCGATTGTCTGTTTCAGGATCTGGTGTGCTT
GGGGGCCGCCTATTGGCTGCCATTGTTCCACCAGGTGTTGACCCCACA

Figure 7. Phylogenetic relationship of the VP1's partial length genomic sequence of spotted hyena (*Crocuta crocuta*) *Sapovirus* strain (original) indicated by a black triangle




Phylogenetic relationship based on the maximum-likelihood (ML) method (T92+G+I model) for a fragment of 351 nucleotides of the VP1 encoding gene describing the relationship between the sapovirus (SaV) strain from spotted hyena determined in this study (marked with a triangle) and other strains from the same genus. Statistical support for nodes was provided by 1000 replicates and the numbers at the nodes indicate the bootstrap values. The scale represents the nucleotide substitutions per site. For each strain are quoted the known genogroup (GI-GXV) based on Oka et al. (2016) proposal, the host species, the origin and the accession number on GenBank.

## 4. DISCUSSION

The epidemiological surveillance of wild animals provides knowledge on the pathogens and diseases in wildlife populations, their geographic distribution, and the appearance of new pathogens in time and space. This information is gathered through the continuous monitoring of wildlife population in the field, and the collection of samples and their analysis (Dufour & La Vieille, 2000; Leighton, 2010)

Long-term surveillance studies like the one reported by Olarte-Castillo *et al*. (2016) are important to know the pathogens circulating in the Serengeti NP, more specifically in the spotted hyena's populations. The present work proposed to sequence the VP1 region of the SaVs found in spotted hyenas in order to identify the existence of different strains within this population over the years, which could explain the recurrence of outbreaks detected during the mentioned study. This objective was not fulfilled since only a partial sequence of that region from one sample was the obtained.

### 4.1. Sapovirus Sequences

Sample I 600 was successfully sequenced in a small fragment of 351 nucleotides. The remaining two samples in the study (I 579 and M 679), despite presenting the virus that was detected by the amplification of a fragment of RdRp encoding region, were not successfully amplified in the VP1 region.

All three samples were collected in the non-outbreak year 2011 (Olarte-Castillo et al., 2016) so that, following the hypothesis that the outbreaks resulted from the emergence of antigenically different strains, it would be expected that all the viruses used in this study presented a similar antigenicity and possibly a genetically similar VP1 region. Besides that, samples I 600 and M 679 had already revealed in the same study a complete genetic homology for a 208 nucleotides' fragment of the viruses RdRp region. Also, samples I 600 and I 579 were collected from hyenas from the same clan and, therefore, close related.

The fact that the same set of primers was unable to amplify the VP1 region's fragment in two of the samples suggests the existence of genetic diversity between the strains in this region. Since both samples not amplified were collected in January 2011 and the sample amplified and sequenced was collected in July of the same year, it can be proposed that occurred a genetic evolution of the virus between January and July. However, this genetic divergence between the strains possibly did not translate into an aa diversity to the point of changing the viruses' antigenicity.

Regarding the I 600 sample, its VP1 amplification was only possible for a small fragment of 351 nucleotides using primers previously described in literature. The impossibility of sequencing the remaining genome is due to the great genetic diversity of this region and consequent lack of knowledge required to design gene-specific primers for this region. Several assays on SaV have resorted to multiple or degenerated primers to achieve the

sequence of RdRp to VP1 region using conventional RT-PCR methods (Jiang et al., 1999; Ludert et al., 2004; Zintz et al., 2005). However, the process of designing a pair of primers to fit an unknown genome as diverse as the SaV is complex and consumes time and resources. Nowadays new and more efficient sequencing technologies exist and have already been used on SaV' studies in order to achieve not only the VP1 region, but also the complete genomic sequence (Oka et al., 2016). The next-generation techniques, based on a specific primer-independent metagenomics sequence approach, allow the sequencing of the nearly complete genome (Li et al., 2011a; Li et al., 2011b). These technologies consist on the fragmentation of the genetic material and the simultaneous sequencing of those fragments. The resulting pool of sequences is later aligned, using bioinformatics techniques, based on a reference genome of the sequenced species or a related organism that has sufficient genetic similarity (Isakov & Shomro, 2011; Illumina, 2017). The missing part of the genome (5' end or both 5' and 3' ends) can be achieved using the 5' rapid amplification of cDNA ends (RACE) (Nakanishi et al., 2011; Oka et al., 2016).

### 4.2. Spotted Hyena' *Sapovirus* and Phylogeny

With a single sequenced fragment of SaV from one spotted hyena it was not possible to estimate the phylogenetic relation between the strains previously detected in this species in Serengeti NP (Olarte-Castillo et al., 2016).

The phylogenetic analysis performed estimates the relation between the sequenced spotted hyena's virus and other SaV from the proposed 15 genogroups (Oka et al., 2016). However, the majority of information on that tree cannot be considered since the bootstrap values that support the clustering of the majority of the nodes is under 70%, the level of statistical significance required to indicate a similarity between strains placed in the same cluster.

# IV. ALPHACORONAVIRUS RECEPTOR IN WILD CARNIVORES

## 1. LITERARTURE REVIEW

### 1.1. *Coronavirus*

The CoVs are a group of viruses included in the subfamily *Coronavirinae* (order *Nidovirales,* family *Coronaviridae). S*ince 2011 they are divided into four genus based on phylogenetic clustering: *Alphacoronavirus*, *Betacoronavirus* (β-CoV), *Deltacoronavirus* and *Gammacoronavirus* (γ-CoV) (Fehr & Perlman, 2015; International Committee on Taxonomy of Viruses [ICTV], 2017).

CoVs' virions are enveloped with a spherical surface of approximately 125 nm diameter and club-shape spike projections that give the appearance of a solar corona to electron microscopy. Within the envelope these viruses have helically symmetrical nucleocapsids with RNA genome (Neuman et al., 2006; Bárcena et al., 2009).

The viral genome is a non-segmented positive sense RNA with approximately 30 kb. Its major gene, which occupies about two-thirds of the genome, is termed replicase and presents two ORFs (rep1a and rep1b) that encode the NSs (NS1-NS16). This gene is followed by the genes that encode the structural and accessory proteins, which occupy the remaining third of the genome in the 3' end (Zhao et al., 2012).

The nucleocapsid (N), envelope (E), membrane (Mb) and spike (S) proteins are the main structural proteins of the virions. N protein is the only present in the nucleocapsid (Fehr et al., 2015). E protein, highly diverse in different CoVs, is present in low quantities in the virions and among its functions facilitates their assembly and release (Godet, L'Haridon, Vautherot & Laude, 1992; Nieto-Torres et al., 2014). Mb protein, in opposition, is the most abundant and is assumed to be responsible for the virions' shape (Fehr et al., 2015). S protein forms the spike structure on virus' surface and is responsible for its attachment to the host receptor. It is, therefore, the primary determinant for the CoV tropism and ability to infect and assumed to influence its host-species specificity. This structure is also subject to positive selection inducing the production of virus-neutralizing antibodies (Tooze, Tooze & Warren, 1984; Delmas, Laude & Agronomique, 1990; Kuo, Godeke, Raamsman, Masters & Rottier, 2000; Gallagher & Buchmeier, 2001; Beniac, Andonov, Grudeski & Booth, 2006; Decaro & Buonavoglia, 2008). A subset of β-CoVs presents a fifth structural protein, hemagglutinin-esterase, which is thought to improve the entry of the virus in the cell, mediated by the S protein, and its spread through the mucosa (Cornelissen et al., 1997).

### 1.1.1. Life cycle

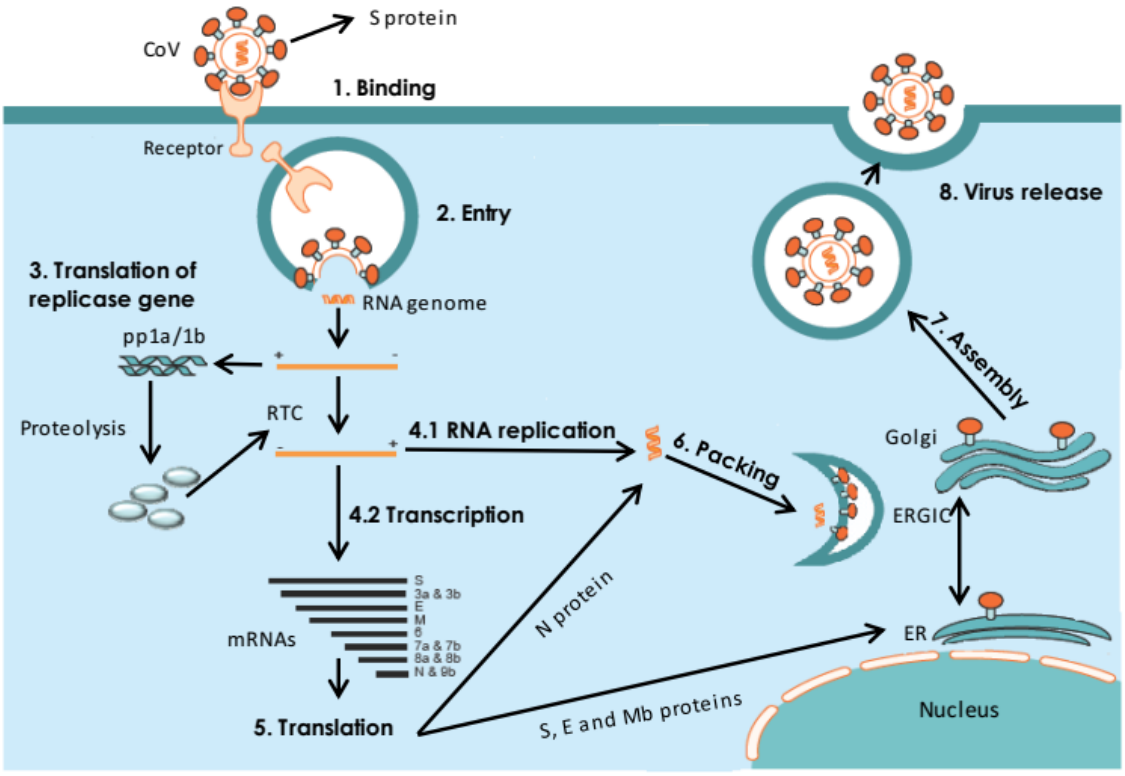
The murine hepatitis virus (MHV), a β-CoV, is the most studied animal CoV and is considered the ideal model for studying the basics of viral replication in tissue culture cells (Fehr et al., 2015).

In order to infect a host cell the CoV binds to a specific cell receptor through the S protein (figure 8). The majority of CoVs enters the host cell through endocytosis, after which the S protein is cleaved and the viral membrane fuses with the endosomal one allowing the viral RNA to enter the cytosol. However, some of these viruses fuse directly with the plasma membrane (Fehr et al., 2015).

In cytosol, viral RNA acts as mRNA and the replicase gene, i.e. rep1a and rep1b, is translated into two polyproteins. These are then cleaved into the NSs by proteases encoded in the viral genome itself (Ziebuhr, Snijder & Gorbalenya, 2000). Several of these NSs are assembled to form the replicase-transcriptase complex where occurs the RNA replication and transcription of sub-genomic RNAs. The RNA resulting from replication will be included in the new virions as their genome, while the sub-genomic RNA will work as mRNAs for the translation of structural and accessory proteins. The structural proteins S, E and Mb are then inserted into the endoplasmic reticulum and follow the secretory pathway into the Golgi complex. Meanwhile, the N protein forms capsids around the positive-sense full-length viral genomes. In the endoplasmic reticulum-Golgi intermediate compartment, within the secretory pathway, the encapsidated viral genomes bud into its membranes which contain the viral structural proteins. The virion assembly is completed and the virions are transported via vesicles to the cell surface where they are released by exocytosis (Tooze et al., 1984; Krijnse-Locker, Ericsson, Rottier & Griffiths, 1994; de Haan & Rottier, 2005; Fehr et al., 2015).

In several of these viruses the S proteins that are not assembled into virions contribute to the virus' spread to adjacent cells. The S proteins are transported to the infected cell's surface and mediate a cell-cell fusion between this and the neighbouring uninfected cells. The resulting multinucleated cells formation allows the virus to spread and infect new cells without being detected or neutralized by virus-specific antibodies (Fehr et al., 2015).

Figure 8. Coronavirus replication model (adapted from Zhu, Liu, Du, Lu & Jiang, 2013)




pp1a/1b – polyproteins 1a and 1b; RTC - replicase-transcriptase complex; mRNAs – messenger RNAs; ER – endoplasmic reticulum; ERGIC - endoplasmic reticulum-Golgi intermediate compartment

### 1.1.2. Pathophysiology

The pathogenesis of CoVs has been actively studied since the 70s because of their impact in livestock and companion animals. However, the importance of this family of viruses was underestimated for a long time, possibly due to the absence of severe diseases that could be undoubtedly attributed to these viruses in humans. In 2003, with the development of the severe acute respiratory syndrome (SARS), caused by a CoV, the *Coronavirinae* subfamily gained new importance (Fehr et al., 2015; Weiss & Navas-Martin, 2005). From that year on these viruses started to be taken into account as emerging pathogens, i.e. agents of infectious disease whose incidence increases following its appearance in a new host population or in the existing population due to long-term changes in its underlying epidemiology (Woolhouse & Dye, 2001; Cleaveland, Haydon & Taylor, 2007).

These viruses can cause disease in both acute and chronic forms in the respiratory tract, intestine or central nervous system depending on the virus tropism, which is determined by its receptor in the host cells (Mcintosh, 1974). Depending on the tissue infected by the virus the transmission can vary from the faecal-oral route to droplet contact or direct contact (Fehr et al., 2015).

CoVs have been described for more than 60 years in humans and animals. As it happens regarding the life cycle, the MHV is the ideal model to study the pathogenesis and immune

response to CoVs since their different variants are able to cause respiratory, enteric, hepatic and neurologic infections in mice, providing a large number of animal models. In addition, the manipulation of this virus requires biosafety level 2 laboratory conditions, contrary to virus from the same family that require biosafety level 3 conditions (Lampert, Sims & Kniazeff, 1973; Weiss et al., 2005)

In humans, CoVs can cause a range of respiratory diseases from common cold, upper respiratory infection, bronchiolitis and pneumonia to severe respiratory tract infections that might be lethal. The α-CoVs, human CoV strain 229E (HCoV-229E) and strain NL63 (HCoV-NL63), and β-CoVs, human CoV strain OC43 (HCoV-OC43) and strain HKU1 (HCoV-HKU1), are endemic in human populations and responsible for mild self-limiting respiratory infections that can become more severe in susceptible groups such as neonates or elderly (Bradburne, Bynoe & Tyrrell, 1967; McIntosh, Becker & Chanock, 1967). The SARS and Middle East respiratory syndrome (MERS) are caused by two β-CoVs, SARS-CoV and MERS-CoV, which crossed the species barrier and infected humans. They are both believed to be originated from bats that transmitted the virus to humans via palm civets (*Paradoxurus hermaphroditus*) and dromedary camels (*Camelus dromedarius*) as intermediate hosts in SARS and MERS, respectively. Both diseases were responsible for severe respiratory infections and several deaths (Lau et al., 2005; Li et al., 2005; Boheemen et al., 2012; Meyer et al., 2014).

Some α-CoVs are known to cause disease in domestic dogs and cats (*Felis cattus*) like canine CoV (CCoV) and feline enteric CoV (FCoV), respectively. Both viruses can be divided in two serotypes (I and II) with equivalent biological properties. Serotypes I and II of CCoV respond differently in culture and the receptor of CCoV I is still unknown (Brownlie, 2017). FCoV type I is the original virus while type II results from the recombination of FCoV type I and CCoV (Benetka et al., 2004). These viruses cause merely mild or asymptomatic enteric infections but the feline infectious peritonitis virus (FIPV), a highly virulent virus with tropism to macrophages that results from a mutation in FCoV during a persistent infection, is capable to cause lethal disease (Perlman & Netland, 2009).

Livestock CoVs have also been identified and are responsible for great economic losses. In pigs, the α-CoVs transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhoea virus (PEDV) cause gastroenteritis in young piglets, with consequent economic losses due to significant morbidity and mortality. In these animals the β-CoV porcine hemagglutinating encephalomyelitis virus causes mainly enteric disease with vomiting and wasting, but can infect the nervous system and cause encephalitis. The cattle can be affected by β-CoV bovine CoV, an agent that can be involved in the winter dysentery in adults and shipping fever disorder in animals of different ages. This virus causes mild to severe respiratory tract infection and diarrhoea leading to weight loss, dehydration, decreased milk production and depression in cattle. It has the ability to spread between a variety of ruminants including elk

(*Cervus canadensis*), deer (Cervidae family) and camels (*Camelus*) crossing the species barrier. Chickens are also affected by a γ-CoV, infectious bronchitis virus, which causes mild to severe respiratory tract infection and in some strains uro-genital tract infection that leads to renal disease. This infection is responsible for significant economic losses due to the reduction of egg production and weight gain (Perlman & Netland, 2009).

### 1.1.3. Alphacoronavirus

The present genus gathers viruses capable to cause disease in humans and animals.

Among the type species of this genus, classified as Alphacoronavirus 1, are the TGEV, CCoV, FCoV and FIPV. Besides the type viruses this genus also holds other animal and human viruses like HCoV-229E, HCoV-NL63 and PEDV. Several α-CoVs have also been identified in wild animals such as mink, bat, spotted hyena or silver-backed jackal. In the latter two species the virus was found in the Serengeti NP within the framework of the research program carried out by the department of Evolutionary Ecology of IZW (Goller et al., 2013; ICTV, 2017).

The great majority of α-CoVs has aminopeptidase N (APN) as host cell receptor (Fehr et al., 2015).

## 1.2. Aminopeptidase N

Aminopeptidase N, also termed alanine aminopeptidase, amino-oligopeptidase, alanyl aminopeptidase (ANPEP), CD13, among others, is a metalloprotease and type II integral membrane protein (MEROPS, 2017).

Metalloproteases are a group of endo- and exopeptidases involved in biological processes from embryonic development and morphogenesis to intestinal absorption of nutrients, or even the metabolism of antibiotics. They are classified into 46 families and these in ten clans based on the metal ion binding motifs and similarities in 3D structure. APN is an exopeptidase belonging to family M1 and clan MA, having HELAH as the aa binding motif that binds zinc ion to it (Look, Ashmun, Shapiro & Peiper, 1989; Noren, Sjöström & Olsen, 1997; Nagase, 2001).

It is classified as a type II integral membrane protein due to its position in the cell membrane, crossing it and having domains both intra- and extracellularly. The latter is protruded from the cell surface in 10,5 nm (Luan & Xu, 2007; Wentworth & Holmes, 2001).

APN is widespread in the organism being part of different epithelial, endothelial and haematopoietic cells. Its presence is more pronounced in the enterocytes of small intestine, the brush border membranes of the kidney and in the liver. Nerve tissues and cells, including the synaptic membranes, and the endothelial cells are also sites where the expression of this enzyme occurs (Luan & Xu, 2007).

### 1.2.1. Protein Synthesis

In humans the locus of the encoding gene ANPEP is 15q25-q26, i.e. is placed in the long arm of chromossome 15, region 2, between bands 5 and 6. In other species this gene is located in different chromosomes, for example in the domestic cat it is present in the subtelomeric chromosome B13 and in the dog in chromosome 3 (Gene, 2004; Luan & Xu, 2007).

The synthesis of a protein from its encoding gene is complex and includes the following steps: (i) transcription, (ii) RNA processing and (ii) translation.

(i) The transcription comprehends the synthesis of a primary transcript or immature/precursor mRNA (pre-mRNA) complementary to a DNA sequence. In eukaryotic cells this sequence will be formed by coding sequences and noncoding sequences interleaved. Both sequences appear in variable number and size in different genes (Watson et al, 2008).

Studies of this gene in humans revealed that its transcription is controlled by two different promoters whose activity is dependent on the tissue where it is expressed. The myeloid promoter is located around 8 kb before the start codon (ATG) and is active in myeloid haematopoietic cells and fibroblasts. The epithelial promoter, that is active in the epithelial cells of the intestine, liver and kidney and in endothelial cells, is present directly before the start codon (Shapiro, Ashmun, Roberts & Look, 1991; Olsen, Kokholm, Norén & Sjöström, 1997). In 1998, Gillis, Pendley & Funkhouser reported the existence of an APN-specific promoter in the alveolar type II epithelial cells of the rat, located 14 kb upstream the start codon. Contrarily to the previous two promoters there is no homology or conservation of this genomic sequence when compared between rat and human (Gillis, Pendley & Funkhouser, 1998).

(ii) In order to originate a RNA sequence capable of being translated into a polypeptide the primary transcript has to be altered through RNA splicing. This process consists in removing noncoding portions of the RNA (introns) leaving just the exons (coding sequences and some noncoding sequences, e.g. 5' and 3' untranslated regions) to form the mature mRNA. The precision is crucial during this process to avoid adding or removing any nucleotide in the joining regions of the exons. A mistake in this process can lead to an incorrect selection of codons downstream to the addition/deletion and the incorporation of the wrong aa into the protein (Watson et al, 2008).

The human APN (hAPN) is known to have 20 exons and the transcripts resulting from epithelial and myeloid promoter differ by 50bp corresponding to the non-coding leader sequence (Olsen et al., 1988; Shapiro et al., 1991; Lerche, Vogel, Shapiro, Norén & Sjöström, 1996; Olsen et al., 1997).

(iii) Once processed, the mRNA travels from the nucleus to the ribosomes in cytosol where translation takes place and the nucleotide sequence is transformed into an aa sequence. In

humans that sequence is constituted by 967 aa (Dybkaer, Kristensen & Pedersen, 2001; Watson et al, 2008).

After translation the amino acid sequence can be subject of modifications (e.g. proteolytic processing, protein folding and glycosylation) that condition the protein structure and determine its ability to interact (Luan & Xu, 2007).

#### 1.2.1.1. RNA Transformation: Alternative Splicing and RNA Editing

After transcription the pre-mRNA is processed into mRNA by RNA splicing, as previously described. However, changes may occur during RNA processing leading to the production of multiple mRNAs, and consequently numerous proteins (isoforms), from a single gene. Those modifications may occur through alternative splicing or RNA editing, two processes that have an important role in the generation of protein diversity within a eukaryote organism explaining their complexity and plasticity even with a limited number of protein encoding genes. Thus, both processes may have an important role in the evolution of gene function (Lynch, 2004; Xing & Lee, 2006; Loya, Van Vactor & Fulga, 2010).

RNA alternative splicing consists in removing sequences from pre-mRNA that are different from the introns removed during RNA splicing. This way, a given pre-mRNA is able to produce more than one mRNA sequence and polypeptide product by being spliced in more than one way. This process can originate from 2 to hundreds or thousands of different polypeptides. It is estimated that 95% of human primary transcripts are spliced in alternative ways and some of the resulting proteins have been involved in the development of some immune diseases and in the metastatic spread of tumours (Soreq et al., 2008; Wang et al., 2008).

Several processes of alternative splicing are described such as mutually exclusive exons, exon extension, exon skipping, intron retention or alternative 5’ or 3’ splicing.

Mutually exclusive exons occur when two exons never appear simultaneously in the different mRNAs resulting from the alternative splicing (figure 9d). That can happen due to the small dimension of an intron between two exons which makes physically impossible to remove only the intron and leads to the removal of the exon downstream (steric hindrance). Another possibility is that the enzymes responsible for cutting a given intron recognize its 5’ sequence but not its 3’ sequence in the pre-mRNA, rather the 3’ of the next intron fit for the enzyme leading to the combinations of major and minor splice sites. Finally it can happen in order to avoid the appearance of a premature stop codon that would happen if both codons featured the mRNA sequence (nonsense-mediated decay) (Watson et al, 2008).

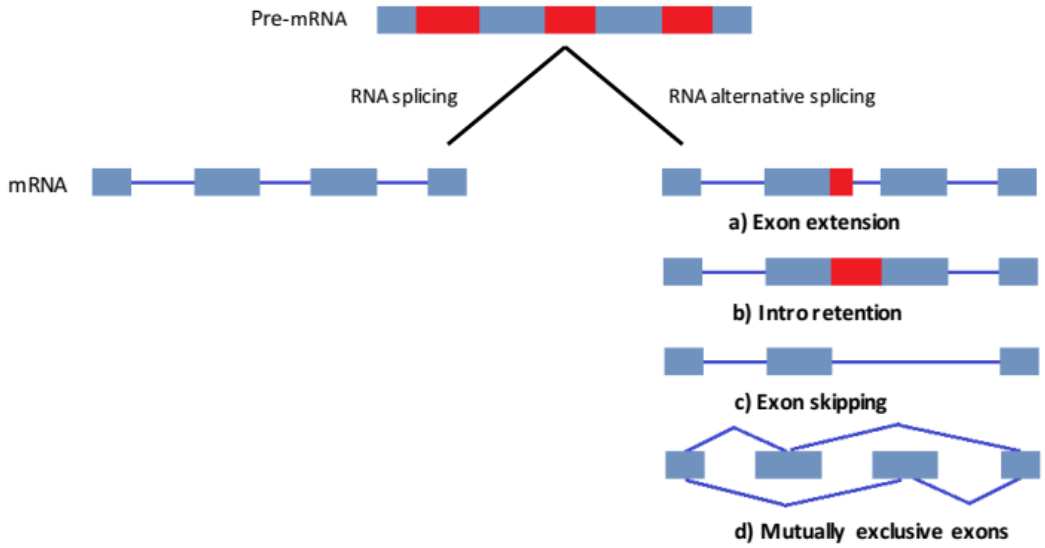
Exon extension (figure 9a) occurs when the mRNA includes part of the intron upstream or downstream to an exon due to selecting a downstream 5’ or an upstream 3’ splice site. Intron retention (figure 9b) occurs when a complete intron appears in the mRNA sequence while in an exon skipping (figure 9c) one complete exon is absent. The latter is the most abundant

type of alternative splicing in mammals (Sugnet, Kent, Ares & Haussler, 2004; Watson et al, 2008).

Alternative 5' or 3' splicing consists in allowing a transcript to include a 5' or 3' exon that normally would not be included in the mRNA (Watson et al, 2008).

Besides the RNA alternative splicing, the RNA editing is another process that can alter the protein prior to translation. In this process single bases are deleted, inserted or changed in specific sites and, consequently, the coding information in the mRNA is altered (Watson et al, 2008).

Figure 9. Patterns of alternative splicing (Original)




APN's encoding gene is no exception and several variants of its mRNA have been identified in several mammals including human, pig, Old world monkeys (e.g. *Chlorocebus sabaeus*, *Macaca nemestrina*), bats (e.g. *Hipposideros aimiger*, *Rinolophus sinicus*) and rodents (e.g. *Cricetulus griseus*, *Chinchilla lanígera*), among others, according to a query for this protein on GenBank (Benson, Karsch-Mizrachi, Lipman, Ostell & Wheeler, 2005).

### 1.2.2. Structure

The one- and two-dimensional structures of APN have been identified but the three-dimensional structure, on which the enzymatic activity depends, is not determined (Luan & Xu, 2007).

The protein molecular weight ranges from 150 to 160 kDa and is formed by two dimers combined by a non-covalent bond, being each of the dimers formed by seven domains (I-VII). Domain I, the cytosolic part of APN, is composed by 9 aa and holds the protein's N terminal which contributes to its anchorage to the cell. The domain that crosses the membrane (II) appears to have a highly conserved alpha-helix structure that varies in

different type II aminopeptidases. All the remaining domains occur outside the cell. Domain III acts similarly to a stalk of the catalytic part of the protease and is followed by domain IV (aa 70 to 252) which, despite not having a significant enzymatic role, is the domain where occurs one of the conserved N-glycosylations. The catalytic part of APN, and consequently the ligand of the zinc-ion part, is formed by domains V and VI (aa 253-580). Finally the domain VII (aa 581-967), a structure with high content of alpha-helix, holds the C-terminal of the protein and is the domain responsible for the noncovalent bond between the dimers (Olsen et al., 1997; Luan & Xu, 2007).

APN is a highly glycosylated enzyme given that at least 20% of its weight is due to the carbohydrates from glycosylation (Luan & Xu, 2007). Glycosylation is a protein's post-translation modification consisting in the enzymatic attachment of glycosyl (donated by a carbohydrate) to it in the endoplasmic reticulum and Golgi complex. This process confers great diversity and is critical for a wide range of biological processes (Jackson, Drummer, Urge, Otvos & Brown, 1994; Wu et al., 1995; Rudd & Dwek, 1997; Wormald et al., 2002).

### 1.2.3. Function

The basic function of APN as an ectoenzyme is to remove the aa from the N-terminal of a diversity of peptides such as the small peptides present in the lumen of small intestine or a variety of biologically active peptides (e.g. enkephalins, angiotensins, neurokinines, cytokines) (Matsas, Stephenson, Hryszko, Kenny & Turner, 1985; Noren et al., 1997; Wentworth et al., 2001; Mina-Osorio, 2008). Through the enzymatic action, APN intervenes in numerous cellular processes like cell cycle control, cell differentiation and motility, inflammation, modulation of immune response, angiogenesis or cellular attachment (Wulfänger et al., 2012). Besides the enzymatic function this protein is also involved in tumorigenesis and viral infection processes.

APN appears over expressed in different solid tumours (e.g. melanoma, liver, renal, pancreas and thyroid tumours, among others) and its activity is elevated in plasma and effusions of cancer patients (Kehlen, Lendeckel, Dralle, Langner & Hoang-vu, 2003; Luan & Xu, 2007). Recent studies have revealed its involvement in the control of cell proliferation, angiogenesis and metastasis in tumours (Mina-Osorio, 2008).

APN is also the confirmed receptor of the majority of α-CoVs strains, namely HCoV-229E, FCoV (types I and II), FIPV, CCoV (types II), TGEV and PEDV (Delmas et al., 1992; Delmas, Gelfi, Sjostrom, Noren & Laude, 1994b; Tresnan, Levis & Holmes, 1996; Fehr et al., 2015). It is also the implicated in the human cytomegalovirus infection and pathogenesis (Luan et al., 2007).

#### 1.2.3.1. APN as *Alphacoronavirus* Receptor

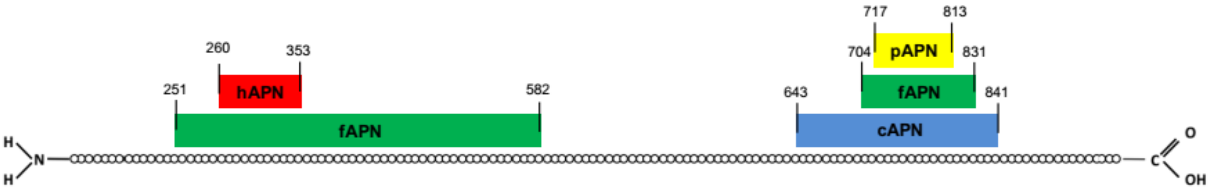
The ability of APN to work as a receptor for α-CoVs has been a study subject aiming the discovery of the specific aa sequence involved in the virus attachment and the possibility of infection to cross the species barrier.

Human, porcine (pAPN), feline (fAPN) and canine (cAPN) APNs have been studied for this purpose. Usually APN receptors are species-specific with hAPN and pAPN working as receptors for HCoV-229E and porcine CoVs (TGEV and PEDV), respectively. Neither of these works as a receptor for an α-CoV that infects other species (Delmas et al., 1994a; Kolb, Maile, Heister & Siddell, 1996). However, *in vitro* studies have demonstrated the capacity of fAPN to act as a receptor not only for feline (FCoV and FIPV), but also human (HCoV-229E), porcine (TGEV) and canine (CCoV type II) CoVs (Tresnan et al., 1996). The same thing was verified for cAPN which is able to act as a receptor for CCoV type II, FCoV type II and TGEV (Benbacer, Kut, Besnardeau, Laude & Delmas, 1997).

The elaboration of studies using APN chimeras allowed the identification of the region within the protein required for the virus attachment and entry in the cell, but the specific aa involved in the process are still to be determined. The protein region determined by these studies differs between different species. In the hAPN the virus receptor activity is located between aa 260 and 353 (figure 10), with special importance for the eight residues (aa 288 to 295) (Luan & Xu, 2007). For the pAPN, in order to TGEV attach to the host cell, the aa between 717 and 813 are required (figure 10) (Delmas et al., 1994a).Chimeric studies on fAPN have revealed that the N-terminal region of the protein is required for HCoV-229E entry while the entry of viruses from cat, pig and dog occur through the interaction with the C-terminal region. Tusell & Holmes (2006) restricted the region between aa 251 and 582 for the entry of HCoV-229E in the cells using fAPN and the region between aa 704 and 831 as the necessary for the entry of TGEV, FIPV and CCoV type II (figure 10). Within the latter region were identified two separate segments (aa 732-746 and aa 764-788) that are essential simultaneously for an effective entry of these viruses in cat cells through fAPN.

In the cAPN Benbacer et al. (1997) also limited the residues necessary for CCoV type II, TGEV and FCoV type II entry to aa 643 – 841 (figure 10).

Figure 10. Diagram of APN's receptor region for alphacoronaviruses (Original)




Above the polypeptide sequence are placed the bands corresponding to the region where the virus is thought to attach in order to enter the cell. Each colour corresponds to a species. In the ends of each band are placed the amino acid numbers that constitute them, according to their species' protein.
hAPN – human APN; fAPN – feline APN; pAPN – porcine APN; cAPN – canine APN.

### 1.3. Aim of the Study

A successful pathogen jump from its existing host to a new host species is dependent on a number of factors, including the latter's susceptibility, which varies greatly. Experimental cross infections (Perlman & Jaenike, 2003; de Vienne, Hood & Giraud, 2009; Longdon, Hadfield, Webster, Obbard & Jiggins, 2011) and reconstructions of host shifts in nature (Davies & Pedersen, 2008; Cooper, Griffin, Franz, Omotayo & Nunn, 2012; Hadfield, Krasnov, Poulin & Nakagawa, 2014; Huang, Bininda-Emonds, Stephens, Gittleman & Altizer, 2014; Waxman, Weinert & Welch, 2014) have demonstrated that pathogens shift preferentially between species phylogenetically closely related. This may happen due to the similarity of the molecular and cellular environment between those species, since a pathogen adapted to interact with the original host's cells in order to infect them, to utilize resources and to avoid or supress the immune system will more successfully infect species with the same or similar environment. Thus, closely related species can be considered more susceptible to infection by a given pathogen (Longdon et al., 2014). However, some pathogens are able to jump between species with great phylogenetic distances (Campisano et al., 2014; Li et al., 2014).

For a virus to be able to enter and infect a cell it is required that its attachment proteins bind to the cell receptor. It is known that some viruses are able to change these proteins and adapt to new receptors enabling the spread of infection into new species (Moya et al., 2004; Allison et al., 2014; Shi, Wu, Zhang, Qi & Gao, 2014). However, if different species present high similarity for the cell receptor of a given virus, it may easily jump and successfully infect different hosts without having to change greatly its attachment region.

In a study conducted between 2003 and 2008 in the centre of Serengeti NP Goller *et al*. (2013) detected the presence of α-CoV in faecal samples from spotted hyenas (suborder Feliformia) and silver-backed jackal (suborder Caniformia) in an area where domestic hosts of CoV were prohibited. The phylogenetic analysis based on the sequence of a fragment of S protein's encoding gene revealed five genetically distinct strains. A strain detected on silver-backed jackal was clustered with CCoV type II strains, like the strains from spotted hyenas

collected in 2006 and 2007. However, the strains from the two species were clustered separated from each other. The remaining two distinct strains from spotted hyenas detected within 2004 were clustered together, apart from the strains previously mentioned, and closer to FCoV type II and TGEV strains. The separate cluster of strains from both wild carnivores reinforced the suggestion that CoV presents a degree of host-specificity in these species and the genetic diversity detected between years suggests a rapid evolution of CoV with the possibility of transmission of strains between species (Jenkins, Rambaut, Pybus & Holmes, 2002; Goller et al., 2013).

The viruses detected in those African wild carnivores were phylogenetically closely related to CoVs know to use APN as receptor.

The aim of my study was to verify the phylogenetic relation between the APN of different wild and domestic carnivores from both Feliformia and Caniformia suborders in order to understand if phylogenetic related species present phylogenetically related APNs. Of particular interest was the aa region known to interact with CoVs' S protein, and my research aimed to determine if the phylogenetic relation of this particular region can indicate the possibility of successful host shifts of CoVs between carnivore host species. My research also aimed to detect the existence of APN isoforms in the different animals. To achieve the aim of this study the APN's mRNA of different wild carnivores, from different tissues where the protein is expressed, were sequenced using conventional RT-PCR methods.

The results obtained in the present study contributed for Ximena A. Olarte-Castillo's doctoral dissertation (2017) and a scientific paper on this subject is being prepared.

## 2. MATHERIAL AND METHODS

### 2.1. Samples

A total of 9 tissue samples from 6 wild carnivore' species from suborders Feliformia and Caniformia were included in the beginning of the study: three samples from leopard (intestine, kidney and liver); two samples from African civet (intestine and kidney); kidney samples from serval, tiger (*Panthera tigris*) and wolf (C*anis lupus*); and one sample of white blood cells from spotted hyena.

The samples from serval, spotted hyena and African civet were collected in Serengeti NP, Tanzania. The leopard's kidney and liver samples and the tiger's kidney sample were obtained from deceased animals housed in the Copenhagen Zoo. The remaining intestine sample from leopard was collected from a leopard in Zoo Berlin. The wolf sample was collected from an animal in Germany (table 6).

After collection, the samples were frozen at -80ºC or preserved in RNAlater at the same temperature or in liquid nitrogen for transportation and storage.

Table 6. Tissue samples from different wild carnivores used in the study, along with their place of collection (Original)

| **Tissue / Species** | **Intestine** | **Kidney** | **Liver** | **White blood cells** |
|---|---|---|---|---|
| Leopard (*Panthera pardus*) | Zoo Berlin | (Copenhagen Zoo) | (Copenhagen Zoo) | |
| African civet (*Civettictis civetta)* | (Serengeti NP) | (Serengeti NP) | | |
| Serval (*Leptailurus serval*) | | (Serengeti NP) | | |
| Tiger (*Panthera tigris*) | | (Copenhagen Zoo) | | |
| Wolf (*Canis lupus*) | | (Germany) | | |
| Spotted hyena (*Crocuta crocuta*) | | | | (Serengeti NP) |

### 2.2. RNA Extraction

The RNA present in the different tissues was extracted using the Direct-zol RNA MiniPrep (Zymo Research, Irvine, California, USA) according to manufacturer's instructions and stored at -80ºC between assays to guarantee its stable preservation.

### 2.3. RT-PCR and Electrophoresis

The amplification of APN's mRNA was performed by RT-PCR using the One *Taq*® One Step RT-PCR kit (New England Biolabs, Ipswich, Massachusetts, USA) following the protocol previously described for Sav amplification: 6,25 µL of reaction mix; 3,75 µL of RNAse free water; 0,5 µL of the enzyme mix; 0,5 µL of each primer [10µM] and 1 µL of the RNA extracted from the tissue, following the instructions on table 2. The combinations of primers and the corresponding annealing temperatures used in the study are described in annex 2. Some of the primers used in this study have already been published while others were modified or newly designed specifically for these samples (table 7).

All RT-PCR products were separated by agarose gel electrophoresis, following the conditions previously described for SaV, and the results visualized under ultraviolet light. When the amplification was successfully achieved and one single amplicon of expected size was observed after electrophoresis, the RT-PCR product was submitted to the purification step and sequenced. When the results of electrophoresis revealed one or more amplicons besides the one of expected size it was attempt to sequence all the amplicons, or at least the one of expected size. In order to separate the amplicons from the same sample and sequence them one of two protocols was applied: DNA extraction from agarose gel or molecular cloning.

Table 7. Primers used on aminopeptidase N amplification and sequencing

| Primer | Sequence (5' to 3') | Sense | Position* | Reference |
|---|---|---|---|---|
| CoVR1F | TGG CCA ARG GNT TCT AYA TTT | Forward | 2 - 22 | Olarte-Castillo, 2017 |
| DT05 | CAT GGC CAA GGG MTT CTA YAT TTC C | Forward | 1 - 24 | Adapted from Tresnan, Levis & Holmes, 1996 |
| AWF | ATC GCT CTG TSC GTG GTS TA | Forward | 79 - 98 | Olarte-Castillo, 2017 |
| CoVR2F | CAT YCA YAG CAA RAR RCT SA | Forward | 381 - 400 | Olarte-Castillo, 2017 |
| CoVR1R | YTA CTC GCT GCG GTA GAA GC | Reverse | 593 - 612 | Olarte-Castillo, 2017 |
| 800f | TAC CTG CTG GCS TAC ATC GT | Forward | 850 - 869 | Olarte-Castillo, 2017 |
| DT03 | CTG GGC YCG GCC YAR TGC MAT | Forward | 924 - 944 | Adapted from Tresnan, Levis & Holmes, 1996 |
| AWR | AAG TCR AGG ATG GGG CCT GT | Reverse | 979 - 998 | Olarte-Castillo, 2017 |
| Gap8r | GTC YTA RAG CAG RGC ACT CTC | Reverse | 1108 - 1128 | Original |
| CoVR3F | AYG ARC TGG CCC AYC AGT | Forward | 1181 - 1198 | Olarte-Castillo, 2017 |
| 1220f | AAY CTG GTG ACC TTG GAR TG | Forward | 1207 - 1226 | Olarte-Castillo, 2017 |
| CoVR2R | YTT CAG CCA RAC YTC ATT CC | Reverse | 1214 - 1233 | Olarte-Castillo, 2017 |
| Gap1f | AYG TGG AGT ACC TAG GTG CTG ACT | Forward | 1265 - 1288 | Original |

* Position on nucleotide sequence based on cAPN (accession number on GenBank: NM001146034)

Table 7 (continuation). Primers used on aminopeptidase N amplification and sequencing

| Primer | Sequence (5' to 3') | Sense | Position* | Reference |
|---|---|---|---|---|
| DT04 | CTC RGC YWY GTC AGC ACC CAG | Reverse | 1283 - 1303 | Adapted from Tresnan et al., 1996 |
| DT04.2 | ASG TGG GTR CRG CYW AGT CAG | Reverse | 1283 - 1303 | Adapted from Tresnan et al., 1996 |
| 1490r | GGT CCT CYG TCA GGA AGR TG | Reverse | 1482 - 1501 | Olarte-Castillo, 2017 |
| Gap3f | GTG MCA TCA TGG ACC GCT GGA TC | Forward | 1631 - 1653 | Original |
| Gap2r | ATG GGY ARY ATC CAC AGG TAM TT | Reverse | 1762 - 1784 | Original |
| CoVR4F | CST CAA YGT GAC RGG CTA YT | Forward | 1893 - 1912 | Olarte-Castillo, 2017 |
| DT07 | GGY GGR GTT GGC AYR AGG GGC A | Forward | 2058 - 2079 | Adapted from Tresnan et al., 1996 |
| DT007.2 | CTG GCG CTG AAC AMC ACY CTC TTC C | Forward | 2056 - 2080 | Adapted from Tresnan et al., 1996 |
| CoVR3R | AGT AYR TCA GGC TGC TCA GG | Reverse | 2118 - 2137 | Olarte-Castillo, 2017 |
| Gap5f | CCT SAG CTA CTT YAR GCT CAT GT | Forward | 2127 - 2149 | Original |
| Gap4r | TAG GAG CAG GCG GTG CTG ACG GCG TTA | Reverse | 2286 - 2312 | Original |
| 2400r | GCY TTG CAG TAC ACG GTY GA | Reverse | 2404 - 2423 | Olarte-Castillo, 2017 |
| Gap7f | CTC CAA CCT CAT CCA GGC VGT GA | Forward | 2718 - 2740 | Original |
| CoVR4R | GTC ACR GCC RGY ATG AGY T | Reverse | 2723 - 2741 | Olarte-Castillo, 2017 |
| CoVR4.2R | GTC ACD SCC TGG ATG AGG TT | Reverse | 2722 - 2741 | Adapted from Olarte-Castillo, 2017 |
| DT08 | CTG GCM CTG YAC AAC ACC CTC TTC C | Reverse | 2733 - 2757 | Adapted from Tresnan et al., 1996 |
| Gap6r | CTT GAA CTG CTC CAG CTG CTG CA | Reverse | 2768 - 2790 | Original |
| DT06 | GGR GGY GTT GGC ARY AGG GGC A | Reverse | 3136 - 3157 | Adapted from Tresnan et al., 1996 |

* Position on nucleotide sequence based on cAPN (accession number on GenBank: NM001146034)

### 2.3.1. DNA Extraction from Agarose Gel

When the electrophoresis result revealed more than one band, corresponding one to the expected size, and the distance between the bands was enough to permit their clear cut from the gel, the process chosen was DNA extraction from agarose gel. This was performed using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) following the corresponding protocol on user manual. After this process the RT-PCR products extracted were sequenced.

### 2.3.2. Molecular Cloning

When the electrophoresis result revealed more than one band, corresponding one to the expected size, and the remaining ones were too close to permit a clear cut from the gel, the process chosen was to clone the amplicons and sequence them separately.

Before actual cloning, a new RT-PCR with 25 µL volume, twice the volume of the previous one, was performed in order to produce a higher quantity of the required DNA.

Following the confirmation of the amplicons' presence by electrophoresis after the new RT-PCR, the remaining volume of the PCR product was purified using QIAquick® PCR Purification kit (Quiagen, Hilden, Germany), according to the protocol "QIAquick PCR Purification kit using a microcentrifuge". The resulting purified DNA was measured by spectrophotometry, using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA), and stored at -20 ºC between uses, to guarantee its stability during the procedure.

The first step of molecular cloning consisted of link the DNA fragments to be cloned into the vector, in an attempt that at least one copy of each amplicon resulting from the RT-PCR was bound to it. In the present work the pDrive Cloning Vector was used for that purpose within the QUIAGEN PCR Cloning kit (Quiagen, Hilden, Germany), following the "QUIAGEN PCR Cloning kit Ligation protocol". The mixture was then stored at -20ºC until cloning. For the following step NEB® Express Competent E. coli (High Efficiency) (New England Biolabs, Ipswich, Massachusetts, USA) were used according to the "High Efficiency Transformation" protocol. This step consisted in integrating the vectors inside cells (one per cell) which were then plated over night at 37ºC on agar plates with ampicillin and X-gal, an analogue of lactose. The vector used in this study has an ampicillin resistance gene which allows the cells transformed by it to grow in this medium. In addition, the pDrive Cloning Vector encodes a protein (LacZ α-peptide) that provides β-galactosidase activity when expressed. Its expression happens when the cell is transformed by the vector but it does not contain a PCR product and, as result, blue colonies grow in the presence of X-gal. When a cell is successfully transformed by the recombinant plasmid white colonies grow. Thus, was possible to distinguish on the plaque by their colour the colonies that corresponded to a successful cloning.

Each white colony was picked from the plates and mixed on a PCR mixture containing 7,5 µL of HotStarTaq Master Mix (Quiagen, Hilden, Germany); 5,75 µL of RNAse free water; and 0,37 µL of M13 forward (-20) primer and the same volume of M13 reverse primer, both primers targeting the pDrive Cloning Vector. The cycling protocol was performed according to the instructions on table 8 and the PCR products were submitted to electrophoresis. The positive PCR products were sequenced.

Table 8. Thermocycling conditions for PCR (QUIAGEN, 2010)

| Cycle step | Temperature | Time | Cycles |
|---|---|---|---|
| Initial activation step | 94ºC | 15 minutes | 1 |
| Denaturation | 94ºC | 30 seconds | 35 |
| Annealing | 50ºC* | 30 seconds | |
| Extension | 72ºC | 1 min/kb | |
| Final extension | 72ºC | 10 | 1 |
| Hold | 4 - 8ºC | ∞ | 1 |

*Annealing temperature for primers M13 forward (-20) and M13 reverse

### 2.4. Sequencing

The DNA fragments to sequence resulting from the original RT-PCR or from the PCR performed within the molecular cloning protocol were submitted to the purification step with exonuclease I and FastAP Thermosensitive Alkaline Phosphatase, according to the protocol on table 4. After this step, all DNA sequences, including the ones resulting from agarose gel extraction, were sequenced following the same protocol described for SaV.

### 2.5. Phylogenetic Analysis

The nucleotide sequences obtained were edited using the Chromas 2.6.4 program (Technelysium Pty Ltd, Australia) and their homology with the APN was verified using the BLAST (Altschul et al., 1997). The phylogenetic analysis was performed on MEGA6 (Tamura et al., 2013).

For the phylogenetic analysis other APN' sequences from different animals were added (table 4). Some of those were retrieved from GenBank with the following accession numbers: human (M22324), domestic pig (KX342854), domestic dog (NM001146034) and domestic cat (U58920). Other sequences from wild carnivores obtained within the framework of Ximena's doctoral dissertation were submitted to GenBank and will be publically available under the following accession numbers when the relevant paper is completed and accepted for publication: cheetah (MF101913), African lion (MF101914), spotted hyena (MF101911), striped hyena (MF101915), brown hyena (MF101908), African wild dog (MF101910), bat-eared fox (MF101906), aardwolf (MF101909), silver-backed jackal (MF101912) and white-tailed mongoose (MF101907) (table 9).

All sequences were aligned using the MUSCLE method (Edgar 2004a,b) based on the sequences' codons. The gaps in the alignments were adjusted manually. The mRNA sequences were translated into protein sequences on MEGA 6 for phylogenetic analysis.

The phylogenetic relationship between the APN amino acid sequence in different species was estimated using the maximum-likelihood (ML) method based on the aa sequence encoded by the smallest partial mRNA of APN obtained in this work (figure 12). The best

model to construct the phylogenetic tree was determined using the BIC, AICc and the InL, which determined the Jones-Taylor-Thorton model (Jones, Taylor & Thornton, 1992) plus gamma distribution (JTT+G) as the most appropriate.

Previously mentioned chimeric experiences using carnivores's APN (fAPN and cAPN) revealed their ability to allow the cell entry of α-CoVs that infect species other than theirs and limited those regions to aa 704-831 (more specifically aa 732-746 and 764 -788) in fAPN and aa 643-841 in cAPN (Tusell & Holmes, 2006; Benbacer et al., 1997). Following these findings, a second phylogenetic tree was constructed based on the largest common aa sequence known to allow the cell entry of α-CoVs from different species through carnivores fAPN and cAPN (figure 13). The estimate of this tree was conducted as the previous one and the Jones-Taylor-Thorton model (Jones et al., 1992) plus gamma distribution (JTT+G) revealed to be the best model.

The statistical support for both analyses was provided by 1000 bootstrap replicates, with the respective percentages quoted on the branches.

Table 9. Species integrated in the phylogenetic analysis (IUCN,2017)

| **Classe** Mammalia | | | |
|---|---|---|---|
| **Order** | **Suborder** | **Family** | **Species or Subspecies – accession number on GenBank** |
| Artiodactyla | Suina | Suidae | Domestic pig (*Sus scrofa domesticus*) - KX342854 |
| Primates | Haplorhini | Hominidae | Human (*Homo sapiens sapiens*) - M22324 |
| Carnivora | Caniformia | Canidae | African wild dog (*Lycaon pictus*) - MF101910<br>Bat-eared fox (*Otocyon megalotis)* - MF101906<br>Black-backed or Silver-backed jackal (*Canis mesomelas*) – MF101912<br>Wolf (*Canis lupus*) - MF101916<br>Domestic dog (*Canis lupus familiaris*) - NM001146034 |
| | Feliformia | Felidae | Cheetah (*Acinonyx jubatus*) - MF101913<br>Leopard (*Panthera pardus*) - MF101918<br>Afican lion (*Panthera leo*) - MF101914<br>Serval (*Leptailurus serval*) - MF101919<br>Domestic cat (*Felis catus*) - U58920<br>Tiger (*Panthera tigris*) - MF101920 |
| | | Hyaenidae | Aardwolf (*Proteles cristatus*) - MF101909<br>Spotted hyena (*Crocuta crocuta*) - MF101911<br>Striped hyena (*Hyaena hyaena*) - MF101915<br>Brown hyena (*Parahyaena brunnea*) - MF101908 |
| | | Viverridae | African civet (*Civettictis civetta)* - MF101917 |
| | | Herpestidae | White-tailed mongoose (*Ichneumia albicauda*) - MF101907 |

## 3. RESULTS

### 3.1. APNs Sequenced

From the original 9 samples, 7 were partially sequenced for APN's mRNA and resulted in continuous sequences. Those corresponded to the intestine samples from African civet and leopard, the kidney samples from African civet, serval, tiger and wolf and the white blood cells from spotted hyena. For the remaining two samples (leopard's kidney and liver) it was possible to partially sequence the mRNA of the APN but they resulted in fragmented nucleotide sequences, reason why they were not include in the later analysis.

The 7 sequences successfully achieved presented different lengths: African civet's APN from intestine (2923 nucleotides), African civet's APN from kidney (2763 nucleotides), leopard's APN (2288 nucleotides), serval's APN (2124 nucleotides), tiger's APN (2296 nucleotides), wolf's APN (3107 nucleotides) and spotted hyena's APN (2213 nucleotides). Some of those were submitted to GenBank, and will be publically available when the related paper is completed and accepted for publication, with the following accession numbers: African civet intestine's APN (MF101917), leopard's APN (MF101918) serval's APN (MF101919), tiger's APN (MF101920) and wolf's APN (MF101916).

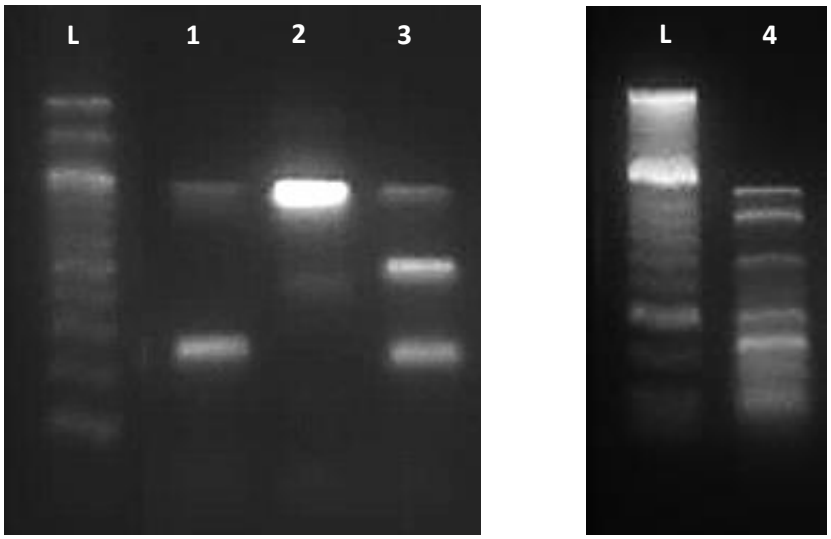
### 3.2. RNA Alternative Splicing

One of the aims of the study was to identify the presence of isoforms of APN as result of alternative splicing. Considering that this process can occur within the region targeted to be amplified by a pair of primers, an electrophoresis can present as result of an amplification by a single pair of primers the amplicon of expected size along with amplicons longer (e.g. because of exon extension or intron retention) and/or shorter (e.g. because exon skipping).

In the current study, after amplification some samples revealed one or more bands besides the one of expected size on gel electrophoresis (figure 11). Using the DNA extraction from agarose gel protocol it was possible, in some cases, to sequence more than one band. However, in those situations the amplicons besides the one of expected size revealed no homology with the existing APN sequences. Using the molecular cloning protocol it was possible to identify some white colonies that when amplified by PCR revealed the presence of an amplicon, which was later sequenced. The sequences achieved revealed homology with the existing APN, however, no more than one amplicon of the same amplified region was obtained for each sample.

Figure 11. Picture of gel electrophoresis' results with more than one band (Original)

a) b)




L – 100 bp DNA Ladder. a) Result of the RT-PCR amplification using the pair of primers CoV4F and CoV4R.2 on samples leopard's kidney (1), civet's intestine (2) and serval's intestine (3). The distanced between the bands allowed the extraction of the band from the gel by its cut. b) Result of the RT-PCR amplification using the pair of primers Gap5f and Gap6r on samples leopard's liver (4). Given the proximity of the bands they didn't allow their clean and direct extraction from the gel.

### 3.3. Phylogenetic Trees

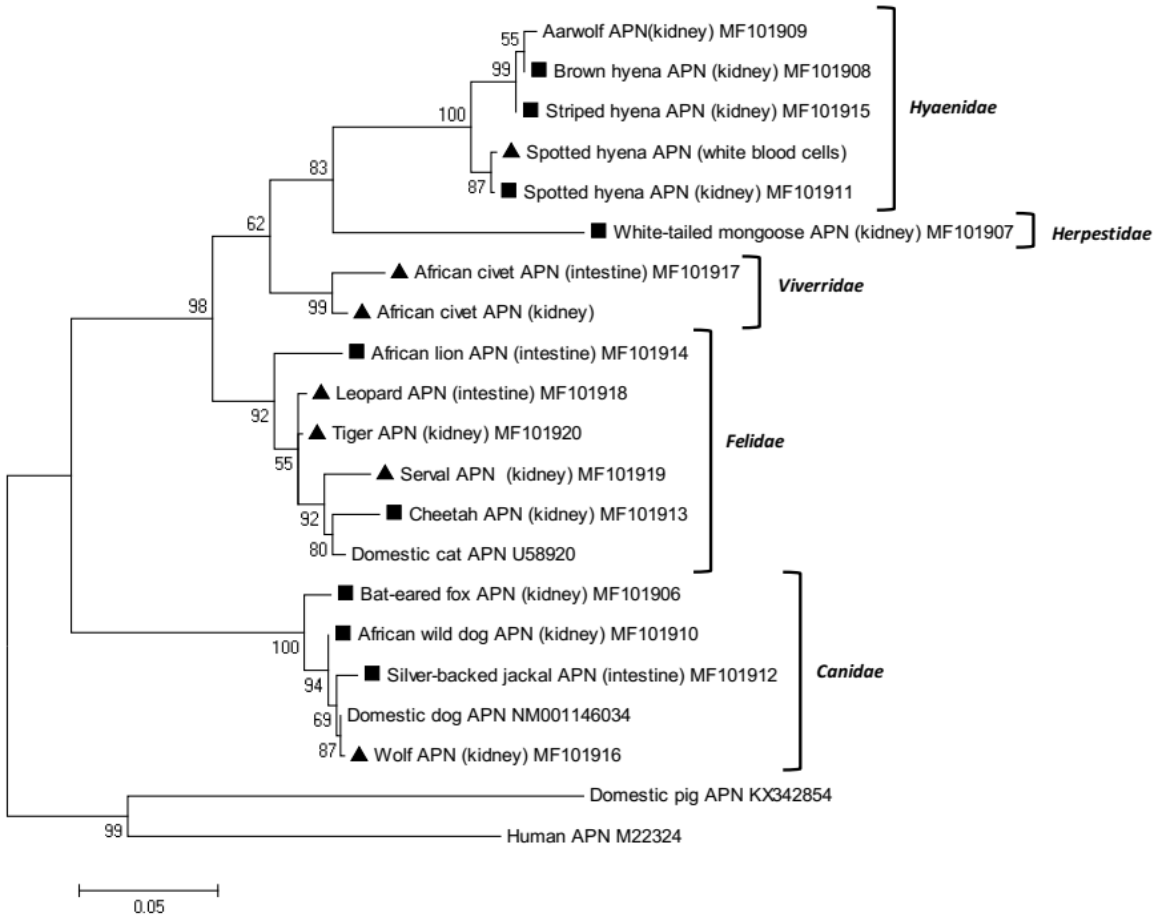
The phylogenetic relationship between the APN of different wild and domestic carnivores, domestic pig and human is present on figure 12. The serval sequence obtained in this work, because it is the shorter one, determined the aa sequence length in which the analysis was based on. Therefore, this phylogenetic tree was constructed based on an aa sequence composed by 644 aa residues, corresponding to aa 329 to 974 from domestic dog's APN or cAPN (accession number on GenBank: NM001146034).

The tree in figure 12 presented a statistical support >70% in the majority of the nodes.

A clear separation was observed between the species from Carnivora order and the sequences from human and domestic pig (99% statistical support). Regarding the carnivore's sequences was verified that the APNs of the species from Canidae family (Caniformia suborder) clustered together with a statistical support of 100%, forming a monophyletic group apart from the remaining sequences of animals from the Feliformia. The APNs from the latter formed a group with a statistical support of 98%. Within this group, the sequences were clustered according to the families of their species of origin with statistical support >70%.

The APNs retrieved from different organs of the same species (spotted hyena's white blood cells and kidney; African civet kidney and intestine) clustered together in monophyletic groups.

Figure 12. Phylogenetic relationship of aminopeptidase N's partial amino acid sequence of wild carnivores (Original)

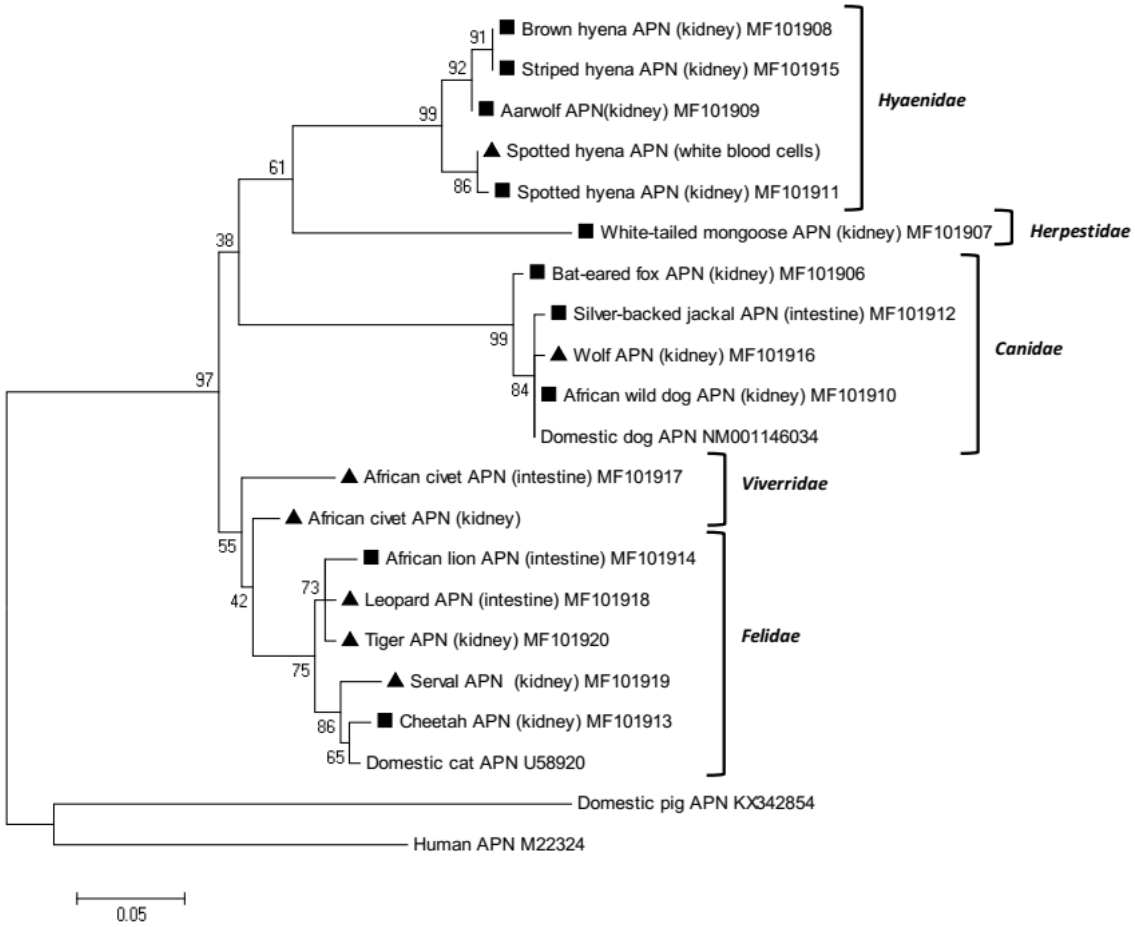



Phylogenetic relationship based on the maximum-likelihood (ML) method (JTT+G) for a fragment of 644 aminoacids of APN. The present tree describes the relationship between the APNs sequenced in the present study (African civet's intestine and kidney, leopard's intestine, serval and tiger's kidney, spotted hyena's white blood cells) and other species from the Carnivora order (domestic dog, domestic cat, APN from the kidney of aardwolf, brown hyena, striped hyena, spotted hyena, white-tailed mongoose, cheetah, bat-eared fox and African wild dog, APN from the intestine of African lion and silver-backed jackal), along with the protein from human and domestic pig. The sequences obtained in the present work are marked with a triangle while the ones obtained by Ximena A. Olarte-Castillo (2017) are marked with a square. Statistical support for nodes was provided by 1000 replicates and the numbers at the nodes indicate the bootstrap values. The scale represents the aminoacids substitutions per site. For each protein are quoted the species, the organ of origin when known and the accession number on GenBank.

The second phylogenetic tree was constructed in order to explore more specifically the phylogenetic relationship between these species' APN in the C-terminal region known to be used for the majority of α-CoV entry in the cell (figure 13). Since species from both Feliformia and Caniformia suborders are present in this analysis the cAPN restricted region, because it is longer than the fAPN and also includes it (figure 10), determined the aa sequence length in

which the analysis was based on. Therefore, this phylogenetic tree was constructed based on an aa sequence composed by 199 aa residues, corresponding to aa 643 – 841 from domestic dog's APN or cAPN (accession number on GenBank: NM001146034).

Figure 13. Phylogenetic relationship of the C-terminal region of aminopeptidase N with known viral receptor function for the majority of alphacoronaviruses of wild carnivores (Original)




Phylogenetic relationship based on the maximum-likelihood (ML) method (JTT+G) for a fragment of 199 aminoacids of APN. The present tree describes the relationship between the APNs sequenced in the present study (African civet's intestine and kidney, leopard's intestine, serval and tiger's kidney, spotted hyena's white blood cells) and other species from the Carnivora order (domestic dog, domestic cat, APN from the kidney of aardwolf, brown hyena, striped hyena, spotted hyena, white-tailed mongoose, cheetah, bat-eared fox and African wild dog, APN from the intestine of African lion and silver-backed jackal), along with the protein from human and domestic pig. The sequences obtained in the present work are marked with a triangle while the ones obtained by Ximena A. Olarte-Castillo (2017) are marked with a square. Statistical support for nodes was provided by 1000 replicates and the numbers at the nodes indicate the bootstrap values. The scale represents the aminoacids substitutions per site. For each protein are quoted the species, the organ of origin when known and the accession number on GenBank.

This tree had a statistical support >70% in the majority of the nodes. It also presented a clear separation between the sequences of domestic pig and human from the remaining of carnivores (97%). A clustering with 38% of statistical support gathered Hyanidae, Herpestidae (Feliformia suborder) and Canidae (Caniformia suborder) families. Those families from Feliformia suborder gathered with a statistical support of 61%, separated from the sequences of Canidae family which formed a group with 99% of statistical support. The remaining sequences from Feliformia suborder clustered together with a statistical support of 55% and within this group according to the families of their species of origin.

## 4. DISCUSSION

A large variety of diseases caused by CoVs occur in both animals and humans and their consequences for morbidity and/or mortality led to a significant push on the research of these viruses in the last half of the 20th century (Perlman & Netland, 2009). The search for understanding the mechanism of infection of these viruses allowed the discovery of the APN protein as the host receptor for the majority of the viruses from *Alphacororonavirus* genus (Delmas et al., 1992, 1994b; Tresnan et al., 1996;Fehr et al., 2015). This protein, present in a diversity of cells, has been widely studied not only in the attempt of determining the region responsible for the virus attachment, but also for its influence in the development of diseases (artritis, cancer, disorders of the central nervous system) when it is abnormally expressed (Nagase, 2001) or even for its use as a tumour market in acute myeloid leukaemia in humans (Fuji et al., 1995).

The present work proposed to verify if the phylogenetic relation between the APN of different wild and domestic carnivores corresponds to the phylogeny of the species, if that is maintained when the phylogenetic analysis is narrowed to the α-CoV attachment region, and detected the existence of isoforms of this protein.

### 4.1. Aminopeptidase N Sequences

From the initial 9 samples used in this work, 7 were successfully sequenced for the APN's mRNA with the use of various primers.

The remaining two samples, leopard's kidney and liver, were only partially sequenced and when assembled resulted in two fragmented sequences. The inability to complete those sequences may be due to the degradation of the RNA which, contrarily to DNA, is highly unstable with the possibility of occurring spontaneously or by enzymatic degradation through ribonucleases present in all cells. The degradation can be prevented if the RNA sample is stored in proper conditions, i.e. -20ºC, -80ºC or liquid nitrogen (Oivanen, Kuusela & Lönnberg, 1998; Seelenfreund et al., 2014). Those conservation conditions were met for the tissue samples after collection and until their storage in the IZW. After the samples' RNA was extracted the resulting RNA was conserved at -80ºC as recommended. Assuming that the storage temperature was guaranteed during transport, RNA degradation may have occurred either before the samples were collected or after RNA extraction. In the first case the RNA degradation may have started due to some degree of decomposition of the organs from which samples were taken. In the second case, despite the fact that when needed for the work the RNA samples were removed from the -80ºC freezer for the shortest time possible and always placed on ice, they may have undergone through some degradation throughout the uses.

### 4.2. Aminopeptidase N Alternative Splicing

During the mRNA amplification several samples revealed more than one band in gel electrophoresis. However, the existence of APN isoforms was not verified.

When it was possible to directly extract the bands from the gel, their sequences revealed that only the band of the expected size for that specific set of primers had homology with the APN. The remaining bands corresponded to sequences from microorganisms, like bacteria, that could be in the sample. The amplification of both APN and microorganisms' fragments through the same set of primers can be explained by the latter low specificity. The specificity of the primers is defined as the frequency with which a mispriming event occurs, i.e. the amplification of something beyond the sequence of interest. In the presence of primers with poor specificity there is a tendency for extra unrelated amplicons to appear in the agarose gel electrophoresis, as it happened in this study (Dieffenbach, Lowe & Dveksler, 1993).

On the other hand, some samples revealed in their electrophoresis results at least one band really close to the one of expected size, which could correspond to two different length mRNAs resulting from alternative splicing. Those PCR products were submitted to the molecular cloning protocol but only one fragment was effectively collected from the plate and sequenced. The absence of colonies with the other DNA fragment could be explained by its loss during the protocol. (i) The lower concentrations of that fragment in the beginning of the process could have determined its presence in small quantity, or even its absence, in the remaining steps of the process. (ii)The incorrect homogenization of the purified DNA solution could have led to the loss of these fragments during pipetting to the step of the PCR product ligation to the vector. (iii) Even if present in normal concentration in the ligation process, the fragment could not have linked to the vector due to an inappropriate ratio between the vector and the PCR products, short time or incorrect temperature of ligation, the presence of nucleases as contamination, over exposition of the PCR products to ultraviolet light, among other factors. (iv) It may also have happened that the cells transformed by the vector with that specific fragment were not pipetted from the tube were the transformation and cell replication occurred and, therefore, were not plated (QIAGEN, 2015). All of these scenarios may explain the absence of more than one fragment per sample per set of primers as a result of molecular cloning and the non-detection of APN isoforms that could have been amplified.

### 4.3. Aminopeptidase N Phylogeny

The first phylogenetic tree constructed in this study based on an aa sequence of 644 residues presents, as expected, the protein's clustering concordant with the phylogenetic relation of their animals of origin with a strong statistical support (figure 12). Thus, the similarity of the molecular environment of phylogenetically related species is demonstrated for the viral receptor (Longdon et al., 2014).

### 4.4. *Alphacoronavirus* Receptor's Phylogeny

The second phylogenetic tree, based on the aa region known to be involved in the α-CoV host cell entry through cAPN and fAPN (figure 13), presents a clear separation between the clustering of the carnivores' protein and the human and domestic pig.

Within the carnivores, the clustering of this specific region of the protein differs from the one observed in the first tree. In this tree was verified separation in the Feliformia suborder with both Felidae and Viverridae families clustering in a monophyletic group apart from other monophyletic group formed by Hyaenidae and Herpestidae. The latter clustered closer to the sequences from Canidae family (from Caniformia suborder). The first monophyletic group presents the expected clustering, similar to the one observed in the first tree, with exception to the fact that the two samples from African civet are not clustered as a monophyletic group. However, their separation in this tree doesn't present a good statistical support (<70%). The second monophyletic group described within the carnivores presents a surprising result by clustering together families from both suborders.

The fact that the α-CoV receptor region of APN in the species from Hyaenidae and Herpestidae is clustered closer to Canidae family than Felidae family suggests the possibility of these animals being more susceptible to be infected by α-CoVs that shift from canids than from felids, assuming that this same region is used for the virus entry in these wild animals' cells.

## V. CONCLUSION

The emergence of infectious diseases in humans is mostly (60,3%) caused by the shift of pathogens from animals and, from these, 71,8% are originated in wildlife (Jones et al., 2008). Wild animals are constantly subjected to infection by a variety of pathogens and their detection, monitoring and control is almost non-existent when compared to that of domestic animals whose surveillance is tightened in order to prevent the spreading of major transmissible diseases through their increasing global circulation (Hueston, 1993).

The research presented in this thesis tried to answer specific questions about the SaVs circulating in the spotted hyena's populations in the Serengeti NP and the genetic similarity between strains in relation to the VP1 region. In essence the research sorts to establish whether or not outbreaks were associated with the occurrence of antigenically different strains. Sequencing a large fragment of highly variable VP1 region of the virus was problematic hence this specific question could not be answered.

The second main question tackled in this thesis was to determine whether the host cell receptor (APN) for α-CoVs on carnivore cells conformed to the known phylogenetic relation between carnivore species. Results revealed a closer relation between Hyaenidae and Herpestidae families to the Canidae family, instead of Felidae family as would be expected from known phylogenetic relationships between these families. The attempt to detect APN isoforms in the several species was not accomplished.

## VI. FUTURE PERSPECTIVES

The results obtained in this thesis also raised several questions to be answered in the future.

To better understand the dynamics of SaV's circulation within a population additional research has to be done regarding not only the virus life cycle, in order to identify its cell receptor and host shift possibility, but also the immune response developed when the host is infected, namely the specific immune response developed during the infection, the average time it takes for the virus to be cleared from the host and no longer excreted and the duration of immune protection. The answer to all of these topics would be of great importance to support the hypothesis of the development of herd immunity in the case of infection of the spotted hyena population in the Serengeti. In order to verify the emergence of new SaV strains antigenically different as the cause of the outbreaks, at least the complete VP1 region of the viruses collected between 2001 and 2012 should be sequenced. Taking into account the great diversity of this region of the genome, and the fact that conventional RT-PCR methods proved to be insufficient to meet that purpose, the use of next-generation sequencing technologies, in which sequencing is independent of the primer specificity, could be a good alternative to meet this goal. However, although these techniques solve the problem of amplification of genetic fragments from the targeted VP1 region of spotted hyena' SaV strains, a new problem would arise. Wild animals are carriers of a great diversity and load of gastrointestinal microorganisms and all these would produce an extremely large number of genetic fragments, leading to the generation of large quantity of sequence data. Without a reference from a relatively similar genetic strain of the targeted SaV it would be extremely difficult to assign each of many thousands of genetic sequences obtained to the targeted SaV strain or to discard unwanted sequences from non-targeted microorganisms. Thus, with the current methods available the sequencing of the SaV VP1 region will remain a very expensive and time consuming process and without any guarantee that the desired results would be obtained.

Regarding the APN, future research should focus on identify the exact regions to which the CoV attaches in order to enter the feline and canine cells and understand if those correspond to same regions in the phylogenetically related animals. The knowledge of the protein's three-dimensional structure would be useful to understand where is spatially the region to which the virus binds and whether its exposure may be affected by any change in post-translational modification, glycosylation. New APN sequences from different organs of the same species should also be performed in order to understand if any significant difference is identified that could explain the virus tropism to specific organs when this protein is present in several tissues. Differences in the protein glycosylation in the different tissues should also be investigated as it can affect the exposure of the region with viral receptor activity to the virus.

## VII. BIBLIOGRAPHY


Allison, A. B., Kohler, D. J., Ortega, A., Hoover, E. A., Grove, D. M., Holmes, E. C. & Parrish, C. R. (2014). Host-specific parvovirus evolution in nature is recapitulated by in vitro adaptation to different carnivore species. *PLoS Pathogens*, 10, 1–17.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research*, 25, 3389–3402.

Bank-Wolf, B.R., Konig, M. & Thiel, H.J. (2010) Zoonotic aspects of infections with noroviruses and sapoviruses. *Veterinary Microbiology*, 140, 204–212.

Bárcena, M., Oostergetel, G. T., Bartelink, W., Faas, F. G. A., Verkleij, A., Rottier, P. J. M., Koster, A. J. & Bosch, B. J. (2009). Cryo-electron tomography of mouse hepatitis virus : insights into the structure of the coronavirion. *Proceedings of the National Academy of Sciences*, 106, 582–587.

Benbacer, L., Kut, E., Besnardeau, L., Laude, H. & Delmas, B. (1997). Interspecies aminopeptidase-N chimeras reveal species-specific receptor recognition by canine coronavirus, feline infectious peritonitis virus, and transmissible gastroenteritis virus. *Journal of Virology*, 71, 734–737.

Benetka, V., Kübber-Heiss, A., Kolodziejek, J., Nowotny, N., Hofmann-Parisot, M. & Möstl, K. (2004). Prevalence of feline coronavirus types I and II in cats with histopathologically verified feline infectious peritonitis. *Veterinary Microbiology*, 99, 31–42.

Beniac, D. R., Andonov, A., Grudeski, E. & Booth, T. F. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology*, 13, 751–752.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2005). GenBank. *Nucleic Acids Research*, 33, D34–D38.

Boheemen, S. Van, Graaf, M. De, Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, M., Osterhaus, A. D. M. E., Haagmans, B. L., Gorbalenya, A. E., Snijder, E. J. & Fouchier, R. A. M. (2012). Genomic characterization of a newly discovered coronavirus associated with Acute Respiratoy Distress Syndrome in humans. *mBio*, 3, 1–9.

Bohm, T. & Höner, O. R. (2015). Crocuta crocuta. *The IUCN RRed List of Threatened Species 2015.*

Boone, R. B., Thirgood, S. J. & Hopcraft, J. G. C. (2006). Serengeti wildebeest migratory patterns modeled from rainfall and new vegetation growth. *Ecology*, 87, 1987–1994.

Bradburne, A. F., Bynoe, M. L. & Tyrrell, D. A. J. (1967). Effects of “new ” human respiratory virus in volunteers. *British Medical Journal*, 3, 767–769.

Brosi, B. J. & Briggs, H. M. (2013). Single pollinator species losses reduce floral fidelity and plant reproductive function. *Proceedings of the National Academy of Sciences*, 110, 13044–13048.

Brownlie, J. (2017). Coronaviridae. In N.J. MacLachlan & E. J. Dubovi (Eds.), *Fenner’s Veterinary Virology*, (5th ed). Academic Press.

Camargo-Sanabria, A. A., Mendoza, E., Guevara, R., Martinez-Ramos, M. & Dirzo, R. (2014). Experimental defaunation of terrestrial mammalian herbivores alters tropical rainforest understorey diversity. *Proceedings of the Royal Society B: Biological Sciences*, 282, 1–6.

Campbell, K. L. I. & Hofer, H. (1995). People and wildlife: spatial dynamics and zones of interaction. In A. R. E. Sinclair & P. Arcese (Eds.) *Serengeti II: Dynamics, Management and Conservation of an Ecosystem*. Chicago, IL, USA: The University of Chicago Press.

Campisano, A., Ometto, L., Compant, S., Pancher, M., Antonielli, L., Yousaf, S., Varotto, C., Anfora, G., Pertot, I., Sessitsch, A. & Rota-Stabelli, O. (2014). Interkingdom transfer of the acne-causing agent, *Propionibacterium acnes*, from human to grapevine. *Molecular Biology and Evolution*, 31, 1059–1065.

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S. & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., Garcia, A., Pringle, R. M. & Palmer, T. M. (2015). Accelerated modern human-induced species losses: entering the sixth mass extinction. *Science Advances*, 1, 1–5.

Ceballos, G., Ehrlich, P. R. & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, E6089–E6096.

Chang, K.-O., Sosnovtsev, S. S., Belliot, G., Wang, Q., Saif, L. J. & Green, K. Y. (2005). Reverse genetics system for porcine enteric calicivirus, a prototype sapovirus in the *Caliciviridae*. *Journal of Virology*, 79, 1409–1416.

Chang, K.-O., Sosnovtsev, S. V., Belliot, G., Kim, Y., Saif, L. J. & Green, K. Y. (2004). Bile acids are essential for porcine enteric calicivirus replication in association with down-regulation of signal transducer and activator of transcription 1. *Proceedings of the National Academy of Sciences*, 101, 8733–8738.

Chiba, S., Sakuma, Y., Kogasaka, R., Akihara, M., Terashima, H., Horino, K. & Nakao, T. (1980). Fecal shedding of vírus in relation to the days of illness in infantile gastroenteritis due to calicivirus. *The Journal of Infectious Diseases*, 142(2), 247-249.

Cleaveland, S., Haydon, D.T. & Taylor, L. (2007). Overviews of pathogen emergence: which pathogens emerge, when and why?. *Current Topics on Microbiology and Immunology*, 315, 85-111.

Cooper, N., Griffin, R., Franz, M., Omotayo, M. & Nunn, C. L. (2012). Phylogenetic host specificity and understanding parasite sharing in primates. *Ecology Letters*, 15, 1370–1377.

Coppola, N., Onorato, L., Minichini, C., Di Caprio, G., Starace, M., Sagnelli, C. & Sagnelli, E. (2015). Clinical significance of hepatitis B surface antigen mutants. *World Journal of Hepatology*, 7, 2729–2739.

Cornelissen, L. A. H. M., Wierda, C. M. H., van der Meer, F. J., Herrewegh, A. A. P. M., Horzinek, M. C., Egberink, H. F. & Groot, R. J. D. E. (1997). Hemagglutinin-esterase, a novel structural protein of torovirus. *Journal of Virology*, 71, 5277–5286.

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife - threats to biodiversity and human health. *Science*, 287, 443–449.

Davies, T. J. & Pedersen, A. B. (2008). Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings of the Royal Society B: Biological Sciences*, 275, 1695–1701.

de Vienne, D. M., Hood, M. E. & Giraud, T. (2009). Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology*, 22, 2532–2541.

Delmas, B., Gelfi, J., Kut, E., Sjöström, H., Noren, O. & Laude, H. (1994a). Determinants essential for the transmissible gastroenteritis virus-receptor interaction reside within a domain of aminopeptidase-N that is distinct from the enzymatic site. *Journal of Virology*, 68, 5216–5224.

Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L. K., Sjöström, H., Norén, O. & Laude, H. (1992). Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature*, 357, 417–20.

Delmas, B., Gelfi, J., Sjostrom, H., Noren, O. & Laude, H. (1994b). Further characterization of aminopeptidase-N as a receptor for coronaviruses. In H. Laude. & J.-F. Vautherot (Eds.), *Coronaviruses. Advances in Experimental Medicine and Biology,Vol.* 342, 293–298. Boston, MA: Springer.

Delmas, B., Laude, H. & Agronomique, R. (1990). Assembly of Coronavirus Spike Protein into Trimers and Its Role in Epitope Expression. *Journal of Virology*, 64, 5367–5375.

Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. (1993). General concepts for PCR primer design. *Genome Research*, 3, S30–S37.

Dufour, B. & La Vieille, S. (2000). Review article. Epidemiological surveillance of infectious diseases in France. *Veterinary Research*, 31, 169–185.

Dybkaer, K., Kristensen, J. S. & Pedersen, F. S. (2001). Single site polymorphisms and alternative splicing of the human CD13 gene - different splicing frequencies among patients with acute myeloid leukaemia and healthy individuals. *British Journal of Haematology*, 112, 691–696.

East, M. L. & Hofer, H. (1991). Loud calling in a female-dominated mammalian society : II. Behavioural contexts and functions of whooping of spotted hyaenas, *Crocuta crocuta*. *Animal Behaviour*, 42, 651–669.

East, M. L. & Hofer, H. (2001). Male spotted hyenas (Crocuta crocuta) queue for status in social groups dominated by females. *Behavioral Ecology*,12, 558–568.

East, M. L., Hofer, H. & Wickler, W. (1993). The erect "penis" is a flag of submission in a female-dominated society: greetings in Serengeti spotted hyenas. *Behavioral Ecology and Sociobiology*, 33, 355–370.

East, M. L., Moestl, K., Benetka, V., Pitra, C., Höner, O. P., Wachter, B. & Hofer, H. (2004). Coronavirus infection of spotted hyenas in the Serengeti ecosystem. *Veterinary Microbiology*, 102, 1–9.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5,113.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32, 1792–1797.

Ehrlich, P. (2014). The case against de-extinction: it's a fascinating but dumb idea. *Yale Environment 360*. (Accessed on 20th September 2017). URL:http://e360.yale.edu/features/the_case_against_de-extinction_its_a_fascinating_but_dumb_idea

Estes, R.D. (1976). The significance of breeding synchrony in the wildebeest. East African. *Wildlife Journal,* 14,135–152.

Farkas, T., Sestak, K., Wei, C. & Jiang, X. (2008). Characterization of a rhesus monkey calicivirus representing a new genus of *Caliciviridae*. *Journal of Virology*, 82, 5408–5416.

Farrell, R. E. J. (2010). Rt-Pcr. In Farrell, R. E. J. (Ed.), *RNA Methodologies: A Laboratory Guide for Isolation and Characterization* (4th ed.), 385–448. Academic Press.

Fehr, A. R. & Perlman, S. (2015). Coronaviruses: an overview of their replication and Pathogenesis. *Methods in Molecular Biology*, 1282, 1–23.

Firth, C., Bhat, M., Firth, M. A., Williams, S. H., Frye, M. J., Simmonds, P., Conte, J. M., Ng, J., Garcia, J., Bhuva, N. P., Lee, B., Che, X., Quan, P.-L. & Lipkin, I. W. (2014). Detection of zoonotic pathogens and characterization of novel viruses carried by commensal *Rattus norvegicus* in New York city. *mBio*, 5, 1–16.

Fleischmann, W. R. Jr. (1996). Viral Genetics. In S. Baron (Ed.), *Medical Microbiology*, (4th ed). Galveston, Texas: University of Texas Medical Branch.

Flynn, W. T. & Saif, L. J. (1988). Serial propagation of porcine enteric calicivirus-like virus in primary porcine kidney cell cultures. *Journal of Clinical Microbiology*, 26, 206–212.

Frame, G. W. (1986). *Carnivore competition and resource use in the Serengeti Ecosystem of Tanzania*. Ph.D. Thesis. Logan, Utah: Utah University State.

Frank, L.G., Glickman, S.E. & Powch, I. (1990). Sexual dimorphism in the spotted hyaena (*Crocuta crocuta*). *Journal of Zoology*, 221(2), 308-313.

Fuji, H., Nakajima, M., Saiki, I., Yoneda, J., Azuma, I. & Tsuruo, T. (1995). Human melanoma invasion and metastasis enhancement by high expression of aminopeptidase N/CD13. *Clinical & Experimental Metastasis*, 13(5), 337-344.

Fullerton, S. W. B., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P. A. & Rohayem, J. (2007). Structural and Functional Characterization of Sapovirus RNA-Dependent RNA Polymerase. *Journal of Virology*, 81, 1858–1871.

Gene (2004) *National Library of Medicine (US), National Center for Biotechnology Information*. (Accessed on 1st November 2017). URL:https://www.ncbi.nlm.nih.gov/gene/?term=ANPEP.

Gillis, L. D., Pendley, D. S. & Funkhouser, J. D. (1998). The major APN transcript of the alveolar type II epithelial cell originates from a unique upstream promoter region. *Biochimica et Biophysica Acta*, 1399, 126–140.

Godet, M., L'Haridon, R., Vautherot, J. & Laude, H. (1992). TGEV Corona Virus ORF4 Encodes a Membrane Protein That Is Incorporated into Virions. *Virology*, 188, 666–675.

Goller, K. V., Fickel, J., Hofer, H., Beier, S. & East, M. L. (2013). Coronavirus genotype diversity and prevalence of infection in wild carnivores in the Serengeti National Park, Tanzania. *Archives of Virology*, 158, 729–734.

Gordis, L. (2014). The epidemiological approach to disease and intervention, *Epidemiology*, (5th ed). Philadelphia, PA: Elsevier Saunders.

Guo, M., Chang, K. O., Hardy, M. E., Zhang, Q., Parwani, A. V & Saif, L. J. (1999). Molecular characterization of a porcine enteric calicivirus genetically related to Sapporo-like human caliciviruses. *Journal of Virology*, 73, 9625–9631.

Guo, M., Evermann, J.F. & Saif, L.J. (2001). Detection and molecular characterization of cultivable caliciviruses from clinically normal mink and enteric caliciviruses associated with diarrhea in mink, *Archives of Virology*, 146 (3), 479-493.

Gutiérrez-Escolano, A. L., Velázquez, F. R., Escobar-Herrera, J., Saucedo, C. L., Torres, J. & Estrada-García, T. (2010). Human caliciviruses detected in mexican children admitted to hospital during 1998–2000 , with severe acute gastroenteritis not due to other enteropathogens. *Journal of Medical Virology*, 82, 632–637.

Hadfield, J. D., Krasnov, B. R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.

Hansman, G. S., Guntapong, R., Pongsuwanna, Y., Natori, K., Katayama, K. & Takeda, N. (2006). Development of an antigen ELISA to detect sapovirus in clinical stool specimens. *Archives of Virology*, 151, 551–561.

Hansman, G. S., Natori, K., Oka, T., Ogawa, S., Tanaka, K., Nagata, N., Ushijima, H., Takeda, N. & Katayama, K. (2005a). Cross-reactivity among sapovirus recombinant capsid proteins. *Archives of Virology*, 150, 21–36.

Hansman, G. S., Natori, K., Ushijima, H., Katayama, K. & Takeda, N. (2005b). Characterization of polyclonal antibodies raised against sapovirus genogroup five virus-like particles. *Archives of Virology*, 150, 1433–1437.

Hansman, G. S., Oka, T., Sakon, N. & Takeda, N. (2007). Antigenic Diversity of Human Sapoviruses. *Emerging Infectious Diseases*, 13, 1519–1525.

Hansman, G. S., Takeda, N., Oka, T., Oseto, M., Hedlund, K. O. & Katayama, K. (2005c). Intergenogroup recombination in sapoviruses. *Emerging Infectious Diseases*, 11, 1916–1920.

Harada, S., Oka, T., Tokuoka, E., Kiyota, N., Nishimura, K., Shimada, Y., Ueno, T., Ikezawa, S., Wakita, T., Wang, Q., Saif, L. J. & Katayama, K. (2012). A confirmation of sapovirus re-infection gastroenteritis cases with different genogroups and genetic shifts in the evolving sapovirus genotypes, 2002-2011. *Archives of Virology*, 157, 1999–2003.

Harada, S., Okada, M., Yahiro, S., Nishimura, K., Matsuo, S., Miyasaka, J., Nakashima, R., Shimada, Y., Ueno, T., Ikezawa, S., Shinozaki, K., Katayama, K., Wakita, T., Takeda, N. & Oka, T. (2009). Surveillance of pathogens in outpatients with gastroenteritis and characterization of Sapovirus strains between 2002 and 2007 in Kumamoto Prefecture, Japan. *Journal of Medical Virology*, 81, 1117–1127.

Hedlund, K. O., Rubilar-Abreu, E. & Svensson, L. (2000). Epidemiology of Calicivirus Infections in Sweden, 1994–1998. *The Journal of Infectious Diseases*, 181, S275–S280.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemestry*, 83, 779–812.

Hofer, H., Benhaiem, S., Golla, W., & East, M. L. (2016). Trade-offs in lactation and milk intake by competing siblings in a fluctuating environment. *Behavioral Ecology*, 27, 1567–1578.

Hofer, H. & East, M. L. (1993a). The commuting system of Serengeti spotted hyenas: how a predator copes with migratory prey. I. Social organization. *Animal Behaviour*, 46, 547–557.

Hofer, H. & East, M. L. (1993b). The commuting system of Serengeti spotted hyenas: how a predator copes with migratory prey. III. Attendance and maternal care. *Animal Behaviour*, 46, 575–589.

Hofer, H. & East, M. L. (2013). Crocuta crocuta, Spotted hyaenas. In J. Kingdon & M. Hoffman (Eds.), The Mammals of Africa, Vol. V: Carnivores, Pangolins, Equids and Rhinoceroses, 273–281. London: Bloomsbury.

Hosmillo, M., Chaudhry, Y., Kim, D.-S., Goodfellow, I. & Cho, K.-O. (2014). Sapovirus translation requires an interaction between VPg and the cap binding protein eIF4E. *Journal of Virology*, 88, 12213–12221.

Huang, S., Bininda-Emonds, O. R. P., Stephens, P. R., Gittleman, J. L. & Altizer, S. (2014). Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *Journal of Animal Ecology*, 83, 671–680.

Hueston, W. D. (1993). Assessment of national systems for the surveillance and monitoring of animal health. *Reveu Scientifique et Technique (International Office of Epizootics)*, 12(4), 1187-1196.

Illumina. (2017). *An Introduction to Next-Generation Sequencing Technology*.

*International Committee on Taxonomy of Viruses (ICTV). (2017). Virus Taxonomy: 2016 Release (Accessed on 16th October 2017). URL:*https://talk.ictvonline.org/taxonomy/.

Isakov, O. & Shomro, N. (2011). Deep Sequencing Data Analysis: Challenges and Solutions. In M. A. Mahdavi (Ed.), *Bioinformatics - Trends and Methodologies,* 655–678. InTech.

IUCN (2017). *The IUCN Red List of Threatened Species. Version 2017-3*. (Accessed on 10th November 2017). URL: http://www.iucnredlist.org.

Iwakiri, A., Ganmyo, H., Yamamoto, S., Otao, K., Mikasa, M., Kizoe, S., Katayama, K., Wakita, T., Takeda, N. & Oka, T. (2009). Quantitative analysis of fecal sapovirus shedding: identification of nucleotide substitutions in the capsid protein during prolonged excretion. *Archives of Virology*, 154, 689–693.

Jackson, D. C., Drummer, H. E., Urge, L., Otvos, L. & Brown, L. (1994). Glycosylation of a synthetic peptide representing a T-cell determinant of Influenza virus hemagglutinin results in loss of recognition by CD4+ T-cell clones. *Virology*, 199, 422–430.

Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. (2002). Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. *Journal of Molecular Evolution*, 54, 156–165.

Jiang, X., Huang, P. W., Zhong, W. M., Farkas, T., Cubitt, D. W. & Matson, D. O. (1999). Design and evaluation of a primer pair that detects both Norwalk- and Sapporo-like caliciviruses by RT-PCR. *Journal of Virological Methods*, 83(1-2), 145-154.

Jones, D.T., Taylor, W.R. & Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in Biosciences*, 8, 275–282.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L. & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451, 990–993.

Katayama, K., Miyoshi, T., Uchino, K., Oka, T., Tanaka, T., Takeda, N. & Hansman, G. S. (2004). Novel recombinant sapovirus. *Emerging Infectious Diseases*, 10, 1874–1876.

Kehlen, A., Lendeckel, U., Dralle, H., Langner, J. & Hoang-vu, C. (2003). Biological Significance of Aminopeptidase N/CD13 in Thyroid Carcinomas. *Cancer Research*, 63, 8500–8506.

Kirkwood, C. D., Clark, R., Bogdanovic-Sakran, N. & Bishop, R. F. (2005). A 5-year study of the prevalence and genetic diversity of human caliciviruses associated with sporadic cases of acute gastroenteritis in young children admitted to hospital in Melbourne, Australia ( 1998 – 2002 ). *Journal of Medical Virology*, 77, 96–101.

Kitajima, M., Oka, T., Haramoto, E., Katayama, H., Takeda, N., Katayama, K. & Ohgaki, S. (2010). Detection and Genetic Analysis of Human Sapoviruses in River Water in Japan. *Applied and Environmental Microbiology*, 76, 2461–2467.

Kolb, A. F., Maile, J., Heister, A. & Siddell, S. G. (1996). Characterization of functional domains in the human coronavirus HCV 229E receptor. *The Journal of General Virology*, 77, 2515–2521.

Krijnse-Locker, Ericsson, M., Rottier, P. J. M. & Griffiths, G. (1994). Characterization of the Budding Compartment of Mouse Hepatitis Virus: Evidence That Transport from the RER to the Golgi Complex Requires only One Vesicular Transport Step. *Journal of Cell Biology*, 124, 55–70.

Kruuk, H. (1972). *The Spotted Hyena: A study of predation and social behaviour*. Chicago: University of Chicago Press.

Kutnjak, D., Elena, S. F. & Ravnikar, M. (2017). Time-Sampled Population Sequencing Reveals the Interplay of Selection and Genetic Drift in Experimental Evolution of Potato Virus Y. *Journal of Virology*, 91.

L'Homme, Y., Brassard, J., Ouardani, M. & Gagné, M. J. (2010). Characterization of novel porcine sapoviruses. *Archives of Virology*, 155, 839–846.

Lampert, P. W., Sims, J. K. & Kniazeff, A. J. (1973). Mechanism of Demyelination in JHM Virus Encephalomyelitis. *Acta Neuropathologica*, 24, 76–85.

Lau, S. K. P., Woo, P. C. Y., Li, K. S. M., Huang, Y., Tsoi, H., Wong, B. H. L., Wong, S. S. Y., Leung, S., Chan, K. & Yuen, K. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences*, 102, 14040–14045.

Lee, L. E., Cebelinski, E. A., Fuller, C., Keene, W. E., Smith, K., Vinjé, J. & Besser, J. M. (2012). Sapovirus outbreaks in long-term care facilities, Oregon and Minnesota, USA, 2002-2009. *Emerging Infectious Diseases*, 18, 873–876.

Lee, R. M., Lessler, J., Lee, R. A., Rudolph, K. E., Reich, N. G., Perl, T. M. & Cummings, D. A. T. (2013). Incubation periods of viral gastroenteritis: A systematic review. *BMC Infectious Diseases*, 13:446, 1–11.

Lerche, C., Vogel, L. K., Shapiro, L.H., Norén, O. & Sjöström, H. (1996). Human aminopeptidase N is encoded by 20 exons. *Mammalian Genome*, 7(9), 712-713.

Li, J. L., Cornman, R. S., Evans, J. D., Pettis, J. S., Zhao, Y., Murphy, C., Peng, W. J., Wu, J., Hamilton, M., Boncristiani Jr, F., Zhou, L., Hammond, J. & Chen, Y. P. (2014). Systemic spread and propagation of a plant-pathogenic virus in European honeybees, *Apis mellifera*. *mBio*, 5, 1–11.

Li, L., Pesavento, P. A., Shan, T., Leutenegger, C. M., Wang, C. & Delwart, E. (2011a). Viruses in diarrhoeic dogs include novel kobuviruses and sapoviruses. *Journal of General Virology*, 92, 2534–2541.

Li, L., Shan, T., Wang, C., Côté, C., Kolman, J., Onions, D., Gulland, F. M. D. & Delwart, E. (2011b). The Fecal Viral Flora of California Sea Lions. *Journal of Virology*, 85, 9909–9917.

Loibooki, M., Hofer, H., Campbell, K. L. I. & East, M. L. (2002). Bushmeat hunting by communities adjacent to the Serengeti National Park, anzania: the importance of livestock ownership and alternative sources of protein and income. *Environmental Conservation*, 29, 391–398.

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J. & Jiggins, F. M. (2014). The Evolution and Genetics of Virus Host Shifts. *PLoS Pathogens*, 10, 1–8.

Longdon, B., Hadfield, J. D., Webster, C. L., Obbard, D. J. & Jiggins, F. M. (2011). Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathogens*, 7, 1–9.

Look, A. T., Ashmun, R. A., Shapiro, L. H. & Peiper, S. C. (1989). Human myeloid plasma membrane glycoprotein CD13 (gp150) is identical to aminopeptidase N. *Journal of Clinical Investigation*, 83, 1299–1307.

Loya, C. M., Van Vactor, D. & Fulga, T. A. (2010). Understanding neuronal connectivity through the post-transcriptional toolkit. *Genes & Development*, 24, 625–635.

Luan, Y. & Xu, W. (2007). The Structure and Main Functions of Aminopeptidase N. *Current Medicinal Chemistry*, 14, 639–647.

Ludert, J. E., Alcalá, A. C. & Liprandi, F. (2004). Primer pair p289-p290 , designed to detect both noroviruses and sapoviruses by reverse transcription-PCR , also detects rotaviruses by cross-reactivity. *Journal of Clinical Microbiology*, 42, 835–836.

Lynch, K. W. (2004). Consequences of regulated pre-mRNA splicing in the immune system. *Nature Reviews Immunology*, 4, 931–940.

MacPhee, R. D. E. & Greenwood, A. D. (2013). Infectious disease, endangerment, and extinction. *International Journal of Evolutionary Biology*, 2013, 1–9.

Madeley, C. R. (1979). Comparison of the features of astroviruses and caliciviruses seen in samples of feces by electron microscopy. *The Journal of Infectious Diseases*, 139(5), 519-523.

Madeley, C. R. & Cosgrove, B. P. (1976). Caliciviruses in Man. *The Lancet*, 307, 199–200.

Martella, V., Lorusso, E., Banyai, K., Decaro, N., Corrente, M., Elia, G., Cavalli, A., Radogna, A., Costantini, V., Saif, L. J., Lavazza, A., Di Trani, L. & Buonavoglia, C. (2008). Identification of a porcine calicivirus related genetically to human sapoviruses. *Journal of Clinical Microbiology*, 46, 1907–1913.

Matsas, R., Stephenson, S. L., Hryszko, J., Kenny, A. J. & Turner, A. J. (1985). The Metabolism of neuropeptidases: Phase separation of synaptic membrane preparations with Triton X-114 reveals the presence of aminopeptidase N. *Biochemical Journal*, 231, 445–449.

Matson, D. O., Estes, M. K., Tanaka, T., Bartlet, A. V & Pickering, L. K. (1990). Asymptomatic human calicivirus infection in a day care center. *The Pediatric Infectious Disease Journal, 9, 190-196*.

Mayo, M. A. (2002). A summary of taxonomic changes recently approved by ICTV. *Archives of Virology*, 147, 1–2.

Mcintosh, K. (1974). Corona viruses : A Comparative Review. In W. Arber, R. Haas, H. Henle, P. H. Hofschneider, N. K. Jerne, P. Koldovský, H. Koprowski, O. Maaløe, R. Rott, H. G. Schweiger, M. Sela, L. Syruček, P. K. Vogt & E. Wecker (Eds.), *Current Topics in Microbiology and Immunology/Ergebnisse der Mikrobiologie und Immunitätsforschung, Vol. 63,* 85–129. Berlin, Heidelberg: Springer.

McIntosh, K., Becker, W. B. & Chanock, R. M. (1967). Growth in suckling-mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proceedings of the National Academy of Sciences*, 58, 2268–2273.

MEROPS (2017). *Summary for peptidase M01.001: aminopeptidase N* (Accessed on 15th October 2017). URL:https://www.ebi.ac.uk/merops/cgi-bin/pepsum?id=M01.001;type=P

Meyer, B., Müller, M. A., Corman, V. M., Reusken, C. B. E. M., Ritz, D., Godeke, G., Lattwein, E., Kallies, S., Siemens, A., van Beek, J., Drexler, J. F., Muth, D., Bosch, B., Wernery, U., Koopmans, M. P. G., Wernery, R. & Drosten, C. (2014). Antibodies against MERS Coronavirus in Dromedaries , United Arab Emirates , 2003 and 2013. *Emerging Infectious Diseases*, 20, 552–559.

Mills, M.G.L. (1990). *Kalahari Hyaenas. The Comparative Behavioural Ecology of two Species*. London: Unwin Hyman.

Mina-Osorio, P. (2008). The moonlighting enzyme CD13: old and new functions to target. *Trends in Molecular Medicine*, 14, 361–371.

Miralles, R., Moya, A. & Elena, S. F. (1999). Effect of poulation patchiness and migration rates on the adaptation and divergence of vesicular stomatitis virus quasispecies populations. *Journal of General Virology*, 80, 2051–2059.

Model, L. & Burnweit, C. A. (2016). Sapovirus Gastroenteritis in Young Children Presenting as Distal Small Bowel Obstruction: A Report of 2 Cases and Literature Review. *Case Reports in Surgery*, 2016, 1–4.

Mombo, I. M., Berthet, N., Bouchier, C., Fair, J. N., Schneider, B. S., Renaud, F., Leroy, E. M. & Rougeron, V. (2014). Characterization of a Genogroup I Sapovirus Isolated from Chimpanzees in the Republic of Congo. *Genome Announcements*, 2, 1-2.

Monica, B., Ramani, S., Banerjee, I., Primrose, B., Iturriza-Gomara, M., Gallimore, C. I., Brown, D. W., Fathima, M., Moses, P. D., Gray, J. J. & Kang, G. (2007). Human caliciviruses in symptomatic and asymptomatic infections in children in Vellore , South India. *Journal of Medical Virology*, 79, 544–551.

Moya, A., Holmes, E. C. & González-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology*, 2, 279–288.

Murchison, E. P., Schulz-Trieglaff, O. B., Ning, Z., Alexandrov, L. B., Bauer, M. J., Fu, B., Hims, M., Ding, Z., Ivakhno, S., Stewart, C., Ng, B. L., Wong, W., Aken, B., White, S., Alsop, A., Becq, J., Bignell, G. R., Cheetham, R. K., Cheng, W., Connor, T. R., Cox, A. J., Feng, Z. P., Gu, Y., Grocock, R. J., Harris, S. R., Khrebtukova, I., Kingsbury, Z., Kowarsky, M., Kreiss, A., Luo, S., Marshall, J., McBride, D. J., Murray, L., Pearse, A. M., Raine, K., Rasolonjatovo, I., Shaw, R., Tedder, P., Tregidgo, C., Vilella, A. J., Wedge, D. C., Woods, G. M., Gormley, N., Humphray, S., Schroth, G., Smith, G., Hall, K., Searle, S. M. J., Carter, N. P., Papenfuss, A. T., Futreal, P. A., Campbell, P. J., Yang, F., Bentley, D. R., Evers, D. J., Stratton, M. R. & Stratton, M. R. (2012). Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell*, 148, 780–791.

Nagase, H. (2001). Metalloproteases. *Current Protocols in Protein Science*, Chapter 21, 1–13.

Nakagomi, O., Nakagomi, T. & Cunliffe, N. A. (2013). Sapovirus. In A.J. Magil, E. T. Ryan, D. Hill & T. Solomon (Eds.), *Hunter's Tropical Medicine and Emerging Infectious Disease* (9th ed.), 289. Elsevier Inc.

Nakanishi, K., Tatsumi, M., Kinoshita-Numata, K., Tsugawa, T., Nakata, S. & Tsutsumi, H. (2011). Full sequence analysis of the original Sapporo virus. *Microbiology and Immunology*, 55, 657–660.

Nakata, S., Chiba, S., Terashima, H., Yokoyama, T. & Nakao, T. (1985). *The Journal of Infectious Diseases*, 152(2), 274-279.

Nakata, S., Estesl, M. K. & Chiba, S. (2005). Detection of human calicivirus antigen and antibody by enzyme-linked immunosorbent assays, 26, 2001–2005.

Nakata, S., Honma, S., Numata, K., Kogawa, K., Ukae, S., Adachi, N., Jiang, X. I., Estes, M. K., Gatheru, Z., Tukei, P. M. & Chiba, S. (1998). Prevalence of human calicivirus infections in Kenya as determined by enzyme immunoassays for three genogroups of the virus. *Journal of Clinical Microbiology*, 36, 3160–3163.

Neuman, B. W., Adair, B. D., Yoshioka, C., Quispe, J. D., Orca, G., Kuhn, P., Milligan, R. A., Yeager, M. & Buchmeier, M. J. (2006). Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *Journal of Virology*, 80, 7918–7928.

New England Biolabs, Inc. (2016). One Taq® One-step RT-PCR kit Instruction Manual.

Nieto-Torres, J. L., DeDiego, M. L., Verdia-Báguena, C., Jimenez-Guardeno, J. M., Regla-Nava, J. A., Fernandez-Delgado, R., Castano-Rodrigez, C., Alcaraz, A., Torres, J., Aguillella, V. M. & Enjuanes, L. (2014). Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathogens*, 10, 1–19.

Nikolin, V. M., Olarte-Castillo, X. A., Osterrieder, N., Hofer, H., Dubovi, E., Mazzoni, C. J., Brunner, E., Goller, K. V., Fyumagwa, R. D., Moehlman, P. D., Thierer, D. & East, M. L. (2016). Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species. *Molecular Ecology*, 26, 1–20.

Nimagen. (2016). BrightDye™ Terminator Cycle Sequencing kit User Manual.

Noren, O., Sjöström, H. & Olsen, J. (1997). Aminopeptidase N. In A.J. Kenny & C. M. Boustead (Eds.) *Cell-Surface Peptidases in Health and Disease,* 175–191. Oxford: BIOS Scientific Publishers.

Numata, K., Hardy, M. E., Nakata, S., Chiba, S. & Estes, M. K. (1997). Molecular characterization of morphologically typical human calicivirus Sapporo. *Archives of Virology*, 142, 1537–1552.

Oivanen, M., Kuusela, S. & Lönnberg, H. (1998). Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by Brønsted acids and bases. *Chemical Reviews*, 98, 961–990.

Oka, T., Katayama, K., Ogawa, S., Hansman, G. S., Kageyama, T., Ushijima, H., Miyamura, T. & Takeda, N. (2005). Proteolytic processing of sapovirus ORF1 polyprotein. *Journal of Virology*, 79, 7283–7290.

Oka, T., Lu, Z., Phan, T., Delwart, E. L., Saif, L. J. & Wang, Q. (2016). Genetic characterization and classification of human and animal sapoviruses. *PLos ONE*, 11, 1–14.

Oka, T., Miyashita, K., Katayama, K., Wakita, T. & Takeda, N. (2009). Distinct genotype and antigenicity among genogroup II sapoviruses. *Microbiology and Immunology*, 53, 417–420.

Oka, T., Wang, Q., Katayama, K. & Saif, J. (2015). Comprehensive review of human sapoviruses. *Clinical Microbiology Reviews*, 28, 32–53.

Oka, T., Yamamoto, M., Katayama, K., Hansman, G. S., Ogawa, S., Miyamura, T. & Takeda, N. (2006). Identification of the cleavage sites of sapovirus open reading frame 1 polyprotein. *Journal of General Virology*, 87, 3329–3338.

Okada, M., Shinozaki, K., Ogawa, T. & Kaiho, I. (2002) *Archives of Virology*, 147, 1445.

Olarte-Castillo, X. A. (2017). *Molecular diversity, and evolution of viruses infecting wild carnivores and the domestic dog in Africa.* Doctoral dissertation. Berlin: Humboldt Universität zu Berlin.

Olarte-Castillo, X. A., Heeger, F., Mazzoni, C. J., Greenwood, A. D., Fyumagwa, R., Moehlman, P. D., Hofer, H. & East, M. L. (2015). Molecular characterization of canine kobuvirus in wild carnivores and the domestic dog in Africa. *Virology*, 477, 89–97.

Olarte-Castillo, X. A., Hofer, H., Goller, K. V, Martella, V., Moehlman, P. D. & East, M. L. (2016). Divergent sapovirus strains and infection prevalence in wild carnivores in the Serengeti Ecosystem : a long-term study. *PLos ONE*, 11, 1–21.

Olsen, J., Cowell, G. M., Kønigshøfer, E., Danielsen, E. M., Møller, J., Laustsen, L., Hansen, O. C., Welinder, K. G., Engberg, J., Hunziker, W., Spiess, M., Sjöström, H. & Norén, O. (1988). Complete amino acid sequence of human intestinal aminopeptidase N as deduced from cloned cDNA. *FEBS Letters*, 238, 307–314.

Olsen, J., Kokholm, K., Norén, O. & Sjöström, H. (1997). Structure and expression of aminopeptidase N. In S. Ansorge & J. Langner (Eds.), *Cellular Peptidases in Immune Functions and Diseases. Advances in experimental Medicine and Biology* Vol. 421, 47–57. Boston, MA: Springer.

Parwani, A. V., Flynn, W. T., Gadfield, K. L. & Saif, L. J. (1991). Serial propagation of porcine enteric calicivirus in a continuous cell line. effect of medium supplementation with intestinal contents or enzymes. *Archives of Virology*, 120, 115–122.

Perlman, S. J. & Jaenike, J. (2003) Infection success in novel hosts: an experimental and phylogenetic study of Drosophia-parasitic nematodes. *Evolution*, 57(3), 544-557.

Perlman, S. & Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenisis. *Nature Reviews Microbiology*, 7, 439–450.

Pfister, T. & Wimmer, E. (2001). Polypeptide p41 of a norwalk-like virus is a nucleic acid-independent nucleoside triphosphatase. *Journal of Virology*, 75, 1611–1619.

Posada, D., Crandall, K. A. & Holmes, E. C. (2002). Recombination in evolutionary genomics. *Annual Review of Genetics*, 36, 75–97.

QUIAGEN. (2010). HotStarTaq® Plus PCR Handbook.

QIAGEN. (2015). QIAGEN ® PCR Cloning Handbook.

Robel, I., Gebhardt, J., Mesters, J. R., Gorbalenya, A., Coutard, B., Canard, B., Hilgenfeld, R. & Rohayem, J. (2008). Functional characterization of the cleavage specificity of the sapovirus chymotrypsin-like protease. *Journal of Virology*, 82, 8085–8093.

Robertson, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society London B: Biology Sciences*, 153, 234–249.

Roelke-Parker, M., Munson, L., Packer, C., Kock, R., Cleaveland, S., Carpenter, M., O ÄôBrien, S., Pospischil, A., Hofmann-Lehmann, R., Lutz, H., Mwamengele, G., Mgasa, M., Machange, G., Summers, B. & Appel, M. (1996). A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature, 389, 441-445*.

Rudd, P. M. & Dwek, R. a. (1997). Glycosylation : Heterogeneity and the 3D structure of proteins. *Critical Reviews in Biochemistry and Molecular Biology*, 32, 1–100.

Russel, P.J. (2010) *iGenetics: A molecular approach* (3rd ed). San Francisco, CA: Pearson Education, Inc.

Saif, L. J., Bohl, E. H., Theil, K. W., Cross, R. F. & House, J. A. (1980). Rotavirus-like, calicivirus-like, and 23-nm virus-like particles associated with diarrhea in young pigs. *Journal of Clinical Microbiology*, 12, 105–11.

Sanjuán, R. & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Sciences*, 73, 4433–4448.

Sano, D., Pérez-Sautu, U., Guix, S., Pintó, R. M., Miura, T., Okabe, S. & Bosch, A. (2011). Quantification and genotyping of human sapoviruses in the Llobregat river catchment, Spain. *Applied and Environmental Microbiology*, 77, 1111–1114.

Scheuer, K. A., Oka, T., Hoet, A. E., Gebreyes, W. A., Molla, B. Z., Saif, L. J. & Wang, Q. (2013). Prevalence of porcine noroviruses, molecular characterization of emerging porcine sapoviruses from finisher swine in the United States, and unified classification scheme for sapoviruses. *Journal of Clinical Microbiology*, 51, 2344–2353.

Schloegel, L. M., Picco, A. M., Kilpatrick, A. M., Davies, A. J., Hyatt, A. D. & Daszak, P. (2009). Magnitude of the US trade in amphibians and presence of *Batrachochytrium dendrobatidis* and ranavirus infection in imported North American bullfrogs (*Rana catesbeiana*). *Biological Conservation*, 142, 1420–1426.

Schotsaert, M. & García-Sastre, A. (2014). Influenza vaccines: A moving interdisciplinary field. *Viruses*, 6, 3809–3826.

Seelenfreund, E., Robinson, W. A., Amato, C. M., Tan, A.-C., Kim, J. & Robinson, S. E. (2014). Long term storage of dry versus frozen RNA for next generation molecular studies. *PLoS ONE*, 9, 1–6.

Serengeti National Park's Official Site (2000). Animals of the Serengeti. (Accessed on 23rd November 2017). URL: http://www.serengeti.org/animals.html

Shapiro, L. H., Ashmun, R. A., Roberts, W. M. & Look, A. T. (1991). Separate promoters control transcription of the human aminopeptidase N gene in myeloid and intestinal epithelial cells. *Journal of Biological Chemistry*, 266, 11999–12007.

Shi, Y., Wu, Y., Zhang, W., Qi, J. & Gao, G. F. (2014). Enabling the "host jump": structural determinants of receptor-binding specificity in influenza A viruses. *Nature Reviews Microbiology*, 12, 822–831.

Sinclair, A.R.E. & Arcese, P.(1995). *Serengeti II: Dynamics, Management and Conservation of an Ecosystem.* Chicago, IL, USA: The University of Chicago Press.

Shivanna, V., Kim, Y. & Chang, K. O. (2014). The crucial role of bile acids in the entry of porcine enteric calicivirus. *Virology* 456–457, 268–278.

Skerratt, L. F., Berger, L., Speare, R., Cashins, S., McDonald, K. R., Phillott, A. D., Hines, H. B. & Kenyon, N. (2007). Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth*, 4, 125–134.

Smith, E. C., Sexton, N. R. & Denison, M. R. (2014). Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. *Annual Review of Virology*, 1, 111–132.

Sodhi, N. S., Brook, B. W. & Bradshaw, C. J. A. (2009). Causes and consequences of species extinctions. In S.A. Levin (Ed) *The Princeton Guide to Ecology,* 514–520. Princeton, NJ, USA: Princeton University Press.

Soreq, L., Gilboa-Geffen, A., Berrih-Aknin, S., Lacoste, P., Darvasi, A., Soreq, E., Bergman, H. & Soreq, H. (2008). Identifying alternative hyper-splicing signatures in MG-thymoma by exon arrays. *PLoS ONE*, 3, 1–18.

Stattersfield, A.J., Crosby, M.J., Long, A.J.& Wege, D.C. (1998). Endemic bird areas of the world: Priorities for biodiversity conservation. *BirdLife International*: Cambridge, UK.

Sugnet, C. W., Kent, W. J., Ares, M. & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium on Biocomputing*, 77, 66–77.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C content biases. *Molecular Biology and Evolution*, 9, 678–687.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

Terashima, H., Chiba, S., Sakuma, Y., Kogasaka, R., Nakata, S., Minami, R., Horino, K. & Nakao, T. (1983). The polypeptide of a human calicivirus. *Archives of Virology*,78,1-7.

Thermo Fisher Scientific. (2016). Product Information Exonuclease I.

Tooze, J., Tooze, S. & Warren, G. (1984). Replication of coronavirus MHV-A59 in sac-cells: determination of the first site of budding of progeny virions. *European Journal of Cell Biology*, 33(2), 281-293.

Tresnan, D. B., Levis, R. & Holmes, K. V. (1996). Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I. *Journal of Virology*, 70, 8669–8674.

Tse, H., Chan, W. M., Li, K. S. M., Lau, S. K. P., Woo, P. C. Y. & Yuen, K. Y. (2012). Discovery and genomic characterization of a novel bat sapovirus with unusual genomic features and phylogenetic position. *PLoS ONE*, 7, 1–11.

Tusell, S. M. & Holmes, K. V. (2006). Molecular interactions of group 1 coronaviruses with feline APN. In Perlman, S. & Holmes, K. V (Eds.), *The Nidoviruses. Advances in Experimental Medicine and Biology* Vol. 581, 289–291. Boston, MA: Springer.

UNESCO (2010). Serengeti Ecosystem Official Map. *UNESCO: World Heritage Centre*. (Accessed on 5th October 2017). URL:http://whc.unesco.org/en/list/156/multiple=1&unique_number=174

Vinjé, J., Deijl, H., van der Heide, R., Lewis, D., Hedlund, K., Svensson, L. & Koopmans, M. P. G. (2000). Molecular detection and epidemiology of Sapporo-like viruses. *Journal of Clinical Microbiology*, 38, 530–536.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456, 470–476.

Wang, Q., Zhang, Z. & Saif, L. J. (2012). Stability of and attachment to lettuce by a culturable porcine sapovirus surrogate for human caliciviruses. *Applied and Environmental Microbiology*, 78, 3932–3940.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M. & Losick, R. (Eds). (2008). *Molecular Biology of the gene* (6th ed). San Francisco, CA: Pearson Education.

Waxman, D., Weinert, L. A. & Welch, J. J. (2014). Inferring host range dynamics from comparative data: the protozoan parasites of new world monkeys. *The American Naturalist*, 184, 65–74.

Weiss, S. R. & Navas-martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiology and Molecular Biology Reviews*, 69, 635–664.

Wentworth, D. E. & Holmes, K. V. (2001). Molecular determinants of species specificity in the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation. *Journal of Virology*, 75, 9741–9752.

Woolhouse, M.E.J. & Dye, C. (2001). Population biology of emerging and re-emerging pathogens. *Philosophical Transactions of the Royal Society of London B*, 356, 981–1106.

World Wide Fund for Nature (WWF). (2016). *Living Planet Report 2016: Risk and resilience in a new era*. *WWF International*. Gland, Switzerland.

Wormald, M. R., Petrescu, A. J., Pao, Y. L., Glithero, A., Elliott, T. & Dwek, R. A. (2002). Conformational studies of oligosaccharides and glycopeptides: complementary of NMR, x-ray crystallography, and molecular modelling. *Chemical Reviews*, 102, 371–386.

Wu, Z., Kayman, S. C., Honnen, W., Revesz, K., Chen, H., Vijh-Warrier, S., Tilley, S. A., Mckeating, J., Shotton, C. & Pinter, A. (1995). Characterization of neutralization epitopes in the V2 region of human immunodeficiency virus type 1 gp120: role of glycosylation in the correct folding of the V1/V2 domain. *Journal of Virology*, 69, 2271–2278.

Wulfänger, J., Schneider, H., Wild, P., Ikenberg, K., Rodolfo, M., Rivoltini, L., Meyer, S., Riemann, D. & Seliger, B. (2012). Promoter methylation of aminopeptidase N/CD13 in malignant melanoma. *Carcinogenesis*, 33, 781–790.

Xing, Y. & Lee, C. (2006). Alternative splicing and RNA selection pressure - evolutionary consequences for eukaryotic genomes. *Nature Reviews Genetics*, 7, 499–509.

Zhao, L., Jha, B. K., Wu, A., Elliott, R., Ziebuhr, J., Gorbalenya, A. E., Silverman, R. H. & Weiss, S. R. (2012). Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host and Microbe*, 11, 607–616.

Zhu, X., Liu, Q., Du, L., Lu, L. & Jiang,S. (2013) Receptor-binding domain as a target for developing SARS vaccines. *Journal of Thoracic Diseases (Emerging infectious respiratory disease & 10th Anniversary of SARS Epidemics),* 5(2), 142-148.

Ziebuhr, J., Snijder, E. J., & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology*, 81, 853–879.

Zintz, C., Bok, K., Parada, E., Barnes-Eley, M., Berke, T., Staat, M. A., Azimi, P., Jiang, X. & Matson, D. O. (2005). Prevalence and genetic characterization of caliciviruses among children hospitalized for acute gastroenteritis in the United States. *Infection, Genetics and Evolution*, 5, 281–290.

**VIII. ANNEX 1 –** Pairs of primers used for the amplification of *Sapovirus* (Original)

| **Primer pair** | **Expected amplicon length** (based on Manchester strain) | **Annealing temperatures tested (ºC)** | **Final annealing temperature (ºC)** | **Samples** (x- PCR executed) | | |
|---|---|---|---|---|---|---|
| | | | | **I 600** | **I 579** | **M 679** |
| Cali2F + Cali2R | 207 bp | 48 | 48 | x | x | x |
| Cali2Rfwd + SV-G1-R | 1004 bp | 48 | 58 | x | x | x |
| | | 53 | | | | |
| | | 58 | | | | |
| Cali2Rfwd + SV-R2 | 1015 bp | 48 | 58 | x | x | x |
| | | 53 | | | | |
| | | 58 | | | | |
| SV-F13 + SV-R13 (nested PCR – first primer pair) | 802 bp | 45 | -------- | x | x | x |
| | | 48 | | | | |
| | | 51 | | | | |
| SV-F14 + SV-R14 (nested PCR – first primer pair) | 802 bp | 45 | -------- | x | x | x |
| | | 48 | | | | |
| | | 51 | | | | |
| SV-F22 + SV-R2 (nested PCR – second primer pair) | 437 bp | 48 | -------- | x | x | x |
| SaV 1245Rfwd + SV-G1-R | 421 bp | 48 | -------- | x | | |
| | | 53 | | | | |
| SaV 1245Rfwd + SV-R2 | 432 bp | 48 | 48 | x | x | x |
| | | 53 | | | | |
| | | 58 | | | | |
| “New F” + SV-G1-R | 518 bp | 48 | -------- | x | | |
| | | 53 | | | | |
| | | 58 | | | | |
| “New F” + SV-R2 | 529 bp | 48 | 48 | x | x | x |

**IX. ANNEX 2 –** Pairs of primers used for the amplification of aminopeptidase N (Original)

| Primer pair | Expected amplicon length | Annealing temperatures tested (ºC) | Final annealing temperature | Samples (x- PCR executed; mc – molecular cloning executed) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Leopard intestine | Leopard kidney | Leopard liver | African civet kidney | African civet intestine | Wolf | Tiger | Spotted hyena | Serval |
| CoVR1F – CoVR1R | 612 bp | 58,2 | 58,2 | x | x | x | x | x | x<br>mc | x | | x |
| CoVR1F – Gap8r | 1127 bp | 49 | -------- | x<br>mc | x<br>mc | | x<br>mc | x<br>mc | x<br>mc | x<br>mc | x<br>mc | x<br>mc |
| | | 50 | | | | | | | | | | |
| | | 52 | | | | | | | | | | |
| | | 53 | | | | | | | | | | |
| | | 54 | | | | | | | | | | |
| | | 55 | | | | | | | | | | |
| | | 56 | | | | | | | | | | |
| | | 58 | | | | | | | | | | |
| | | 60 | | | | | | | | | | |
| | | 62 | | | | | | | | | | |
| | | 64 | | | | | | | | | | |
| DT05 – DT04 | 1303 bp | 64 | 64 | | x | x | x | x | | | | x |
| DT05 – DT04.2 | 1303 bp | 58 | 64 | x | | | | x | x | | | |
| | | 60 | | | | | | | | | | |
| | | 61 | | | | | | | | | | |
| | | 64 | | | | | | | | | | |
| Awf - Awr | 920 bp | 49 | 49 | x | x | x | | | x | x | x | x |
| Awf – Gap8r | 1050 bp | 49 | -------- | x | | | | | | | | |
| | | 50 | | | | | | | | | | |
| | | 52 | | | | | | | | | | |
| | | 54 | | | | | | | | | | |

| Primer pair | Expected amplicon length | Annealing temperatures tested (ºC) | Annealing temperature | Samples (x- PCR executed; mc – molecular cloning executed) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Leopard intestine | Leopard kidney | Leopard liver | African civet kidney | African civet intestine | Wolf | Tiger | Spotted hyena | Serval |
| CoVR2F – CoVR2R | 853 bp | 50 | 50 | x | x | x | x | x | x | x | x | x |
| | | 51 | | | | | | | | | | |
| | | 53 | | | | | | | | | | |
| | | 54 | | | | | | | | | | |
| CoVR2F – Gap8r | 748 bp | 53 | -------- | | | | | | x | x | | |
| | | 54 | | | | | | | | | | |
| | | 55 | | | | | | | | | | |
| | | 56 | | | | | | | | | | |
| | | 58 | | | | | | | | | | |
| | | 60 | | | | | | | | | | |
| | | 62 | | | | | | | | | | |
| | | 64 | | | | | | | | | | |
| 800f – 1490r | 652 bp | 49 | 49 | x | x | x | x | x | | x | x | x |
| DT03 – Gap2r | 861 bp | 52 | -------- | | | | | | | | | x |
| | | 54 | | | | | | | | | | |
| | | 56 | | | | | | | | | | |
| | | 58 | | | | | | | | | | |
| | | 60 | | | | | | | | | | |
| | | 62 | | | | | | | | | | |
| DT03 – DT08 | 1834 bp | 64 | 64 | x | x | x | x | x | | x | x | x |
| CoVR3F – CoVR3R | 957 bp | 61 | 61 | x | x | x | | x | x | x | x | x |
| 1220f – 2400r | 295 bp | 49 | 49 | x | | | | x | | x | x | x |

| Primer pair | Expected amplicon length | Annealing temperatures tested (ºC) | Annealing temperature | Samples (x- PCR executed; mc – molecular cloning executed) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Leopard intestine | Leopard kidney | Leopard liver | African civet kidney | African civet intestine | Wolf | Tiger | Spotted hyena | Serval |
| Gap1f – Gap2r | 520 bp | 50 | 51 | | x | x | | | | | | x |
| | | 51 | | | | | | | | | | |
| | | 52 | | | | | | | | | | |
| | | 55 | | | | | | | | | | |
| | | 56 | | | | | | | | | | |
| | | 57 | | | | | | | | | | |
| Gap3f – Gap4r | 682 bp | 60 | 62 | x | x | x | | x | x | x | x | x |
| | | 62 | | | | | | | | | | |
| | | 64 | | | | | | | | | | |
| CoVR4F – CoVR4R | 845 bp | 61 | 61 | x | | | x | x | x | | x | x |
| CoVR4F – CoVR4.2 R | 849 bp | 61 | 61 | x mc | x mc | x mc | x mc | x mc | x mc | x | x mc | x mc |
| DT07 – DT06 | 1100 bp | 64 | 64 | x | x | x | x | x | | x | | |
| DT07.2 – DT06 | 1100 bp | 64 | 64 | x | x | x | | x | x | x | x | x |
| Gap5f – Gap6r | 664 bp | 50 | 51 | x mc | x mc | x | x | x mc | x mc | x mc | x mc | x mc |
| | | 51 | | | | | | | | | | |
| | | 52 | | | | | | | | | | |
| | | 55 | | | | | | | | | | |
| | | 56 | | | | | | | | | | |
| | | 57 | | | | | | | | | | |
| Gap7f – DT06 | 440 bp | 60 | 62 | x | x | x | | x | x | x | x | x |
| | | 62 | | | | | | | | | | |
| | | 64 | | | | | | | | | | |